

*Paul S. Davies is the Director of the Division of Policy Evaluation, Office of Research, Evaluation, and Statistics, Office of Policy, Social Security Administration. Kalman Rupp is an economist with the Division of Policy Evaluation.*

## ***An Overview of the National Survey of SSI Children and Families and Related Products***

*by Paul S. Davies and Kalman Rupp*

### ***Summary***

The National Survey of SSI Children and Families (NSCF) is the first nationally representative survey since 1978 of noninstitutionalized children and young adults who currently receive or formerly received Supplemental Security Income (SSI). Over 8,500 interviews were completed between July 2001 and June 2002. The primary objective of the NSCF is to provide data to support research and policy evaluation on the current cross section of children (ages 0 to 17) and young adults (ages 18 to 23) receiving SSI. Following that objective, the survey was designed to answer questions such as those presented below.

- What are the general characteristics of children and young adults receiving SSI and their families?
- What are the patterns of access to and utilization of health care among children and young adults receiving SSI?
- What services are utilized by children and young adults receiving SSI?
- What are the unmet health care and service needs of children and young adults receiving SSI?
- What costs are associated with caring for a disabled child?

- What is the impact on the family of having a disabled child?
- What is the status of young adults with disabilities as they make the transition to adulthood?
- How well are they prepared for that transition?

In addition, the NSCF questionnaire and sample were designed to be comprehensive enough and large enough to address numerous additional policy issues as they emerge. The NSCF fills a gap in the data available to policy analysts by addressing a wide range of topics that cannot be addressed with SSI administrative data and by providing a large sample in contrast to major national survey databases that cover this target population fairly sparsely. A companion article to this overview describes general characteristics of SSI beneficiary children and their families (see Rupp and others 2005/2006, pages 21–48 of this issue). Other topics being examined include disability-related expenditures for SSI children and young adults and labor force participation of the parents of SSI children.

The NSCF data files are accompanied by a detailed User's Manual, which includes a detailed codebook and information about the NSCF sample design, questionnaire design and content, data

collection procedures, variable construction, editing, and variance estimation procedures. In order to facilitate research, the Social Security Administration published the NSCF Public-Use File and survey documentation on its Web site. These products are available at <http://www.socialsecurity.gov/disabilityresearch/nscf.htm>. The NSCF is an outstanding tool for conducting research and policy analysis regarding children and young adults receiving SSI.

## ***Introduction***

---

The National Survey of SSI Children and Families (NSCF) is a nationally representative survey of noninstitutionalized children and young adults who are current or former recipients of Supplemental Security Income (SSI). The primary objective of the NSCF is to provide data to support research and policy evaluation on the cross section of children aged 0–17 and young adults aged 18–23 who were receiving SSI in December 2000.<sup>1</sup> Two groups were excluded from the sample universe: recipients in Alaska, Hawaii, and the Northern Mariana Islands, for logistical and cost considerations; and the almost 13,000 children and young adults receiving SSI, or 1.5 percent of the childhood caseload in December 2000, who lived in a Medicaid institution.

The survey yielded 8,535 completed survey interviews.<sup>2</sup> The interviews provide detailed data on topics including health, functional limitations, and disability status; health care utilization; service utilization, perceived needs, and expenses; education and training; employment and earnings; family income and assets; and health insurance. The data were collected between July 2001 and June 2002 through telephone and in-person interviews. Interviews were conducted in English or Spanish. The Social Security Administration (SSA), in collaboration with Mathematica Policy Research, Inc., designed the NSCF, and Mathematica collected the data and prepared the data files and documentation.

The NSCF fills a gap in the data available to policy analysts for two reasons. First, it addresses a wide range of topics that cannot be addressed with SSI administrative data. Second, it provides a large sample in contrast to major national survey databases that cover this target population fairly sparsely (see Ireys and others (2004) for detailed comparisons). In order to facilitate research, SSA published the NSCF Public-Use file and survey documentation at SSA's Web site. These products are available at <http://www.socialsecurity.gov/disabilityresearch/nscf.htm>.

The SSI program was designed to protect the disabled and elderly from extreme poverty. When the original legislation was drafted, Congress heavily debated

whether to include children in the SSI program. If the intent was to cover out-of-pocket expenses associated with the child's disability, perhaps Medicaid coverage would be sufficient and cash benefits unnecessary. If families of disabled children incur other expenses, such as lost wages of a parent who stays at home to care for the disabled child, then cash benefits in addition to Medicaid coverage would be justified.

In the end, Congress decided to include children in the program and to provide cash benefits. The first SSI payments were issued in January 1974. In most states, Medicaid coverage for children receiving SSI is either automatic or very closely related to the SSI eligibility criteria. Although initially a small portion of the overall SSI program, the childhood rolls have grown steadily over the years. In December 1974, only 70,900 children under the age of 18 received SSI payments, representing 3.8 percent of the total SSI caseload; by December 2005, that number had grown to 1,036,498, or 14.6 percent of the SSI caseload.

The NSCF is the first national survey since 1978 of noninstitutionalized children and young adults receiving SSI. General social surveys such as the Survey of Income and Program Participation and the Current Population Survey typically do not contain enough observations to support detailed analyses of this population. SSA administrative records cover all SSI recipients and contain detailed programmatic information, but they lack data on the demographic and economic characteristics of SSI recipients and their families. Some researchers have conducted case studies and focus groups with children and young adults receiving SSI and their families, but the samples are small and not nationally representative (see, for example, Rogowski and others 2002; Lazear and Worthington 2002). The NSCF fills this critical data gap by combining current and credible survey data on the health and well-being of a nationally representative sample of the target population with current and historical administrative records on program status and receipt of SSI payments.

The NSCF data files are accompanied by a detailed User's Manual, which includes a detailed codebook and information about the NSCF sample design, questionnaire design and content, data collection procedures, variable construction, editing, and variance estimation procedures. The NSCF has the potential to be the premier source of data for policymaking, program planning, and general research on children and young adults receiving SSI.

This article

- presents the major groups of interest, sample sizes, and interview completion rates;
- discusses the design and content of the questionnaire;

- describes the procedures for collecting data;
- provides an overview of the procedures for imputing values for missing data, developing sampling weights, and estimating variance; and
- introduces the major NSCF data products.

### ***Groups of Interest, Sample Sizes, and Completion Rates***

The overall sample design of the NSCF provides a framework for analyzing a wide range of policy issues of current and future interest. During the design phase, SSA identified two major groups for analysis—children and young adults receiving SSI in December 2000 and children and young adults receiving SSI in December 1996—and, within these two groups, several subgroups of interest. The design also included a group of people not receiving SSI as a potential tool for making various comparisons with the two major analysis groups. The groups were identified on the basis of programmatic variables defining SSI receipt in SSA’s administrative records. The groups were reorganized into mutually exclusive sampling strata, and a complicated algorithm was developed to optimize the sample allocation to meet SSA’s analytic needs subject to a number of precision and cost constraints. The major analysis groups and important subgroups are described below, followed by a discussion of sample sizes and completion rates. Potter (2001) and Potter and Diaz-Tena (2003) provide additional details.

#### ***Children and Young Adults Receiving SSI in December 2000***

The first major analysis group includes children and young adults identified in SSA administrative records as SSI recipients in December 2000. This group provides an excellent basis for addressing issues about the current caseload that are of interest to policymakers: the group reflects a fairly recent point in time, and no major changes in program rules affecting child and young adult recipients have occurred since December 2000.

Because of potential differences in the way various program features affect subgroups of interest to policymakers, the December 2000 cross section was designed to provide a reasonable representation of children and young adults receiving SSI by key variables such as age group, sex, and type of impairment. For example, comparisons by sex will be important, particularly in light of the contrast between the substantial overrepresentation of males among children receiving SSI compared with that of females among elderly beneficiaries. The type of impairment—specifically, mental versus

physical impairment—is also of great interest. The distribution of children receiving SSI by age group (0–5, 6–12, 13–17) will play a role in some analyses, for example, when looking at special education, parental employment, or impairment differences among children awarded benefits at different ages. Child-to-adult transition issues are particularly important. Therefore, several subgroups were defined within each major analysis group to ensure that a sufficient number of sample cases would be available for comparative analyses.

To reduce the variability of the sample, these subgroups were proportionately represented through a sequential selection procedure rather than through explicit stratification by these subgroup characteristics. Overall, it was necessary to oversample 17- to 18-year-olds in December 2000 and recipients with a mental disorder. The analysis group of December 2000 recipients can be followed using administrative records to track important outcomes such as SSI participation, receipt of benefits from the Social Security Disability Insurance program, earnings, and mortality.

#### ***Children and Young Adults Receiving SSI in December 1996***

The second major analysis group includes children and young adults identified in SSA administrative records as SSI recipients in December 1996. By including this subgroup, the survey provides a snapshot of the December 1996 caseload at the time of the survey in 2001–2002—about 5 to 6 years later. The survey also includes many characteristics not available from administrative records. Important policy questions relate to longer-term outcomes of SSI program participation, and the inclusion of this analysis group provides a limited longitudinal perspective on the dynamics of the SSI caseload. For example, this group allows researchers to track the experiences of the subset of the child caseload that was potentially affected by the 1996 welfare reform legislation. It may also facilitate other analyses of longer-term outcomes, including comparisons with the sample of SSI children who were not affected by that legislation. For studies regarding welfare reform, it is desirable to examine childhood redeterminations separately from age-18 redeterminations. Therefore, children aged 17–18 in December 1996 and those subject to childhood redeterminations were oversampled in the second analysis group.

#### ***Other Groups***

In addition, SSA identified two groups—former SSI recipients and applicants who had been denied—as potentially important comparison groups for certain analyses. Former SSI recipients are defined as children

who had applied for SSI benefits in 1992 or later and were not receiving benefits at the time of the survey or at welfare reform (1996) but who had received benefits for at least 1 month since 1992. Denied applicants are children who had applied for SSI benefits in 1992 or later but had never received benefits between 1992 and the time of the survey. The inclusion of these groups provides an opportunity for policy analyses that require information on children who had some contact with the SSI program in the past but did not receive benefits in either December 2000 or December 1996. For example, denied applicants can be used as a comparison group for SSI awardees.

### Sample Size and Completion Rates

The unweighted and weighted sample sizes of the major groups described above can be seen in Table 1. A total of 5,006 interviews were completed with nonincarcerated children and young adults who were on the SSI rolls in December 2000; 5,033 interviews were completed with those on the SSI rolls in December 1996. The auxiliary group of former recipients (other than those on the rolls in December 1996) and denied applicants contains 1,767 completed interviews. Characteristics of the sample of children and young adults receiving SSI in December 2000 and the sample of children and young adults receiving SSI in December 1996—their age and sex, and their main health condition and general health status as reported by the respondent—are shown in Table 2.

Mathematica Policy Research, Inc., calculated survey completion rates. The unweighted completion rate is the

number of completed interviews and ineligible sample members divided by the number of attempted interviews.<sup>3</sup> The weighted completion rate represents the proportion of the estimated NSCF study universe with a completed interview. The overall weighted completion rate for the NSCF was 74.4 percent. The completion rate was highest for those who were SSI recipients at the time of the survey, largely because they were more likely to be located. The weighted completion rate for SSI recipients in December 2000 was 84.1 percent, compared with 76.8 percent for recipients in December 1996. Completion rates were lower for those not receiving SSI in December 2000 (71.1 percent) and those aged 17–18 in December 1996 (70.0 percent) (authors' calculations based on Potter and Diaz-Tena 2003, 32).

### Design and Content of the Questionnaire

Two versions of the NSCF questionnaire were designed—one for children and one for young adults. For survey implementation, the two versions were combined into a single, computerized instrument with fairly complex skip patterns. Although the majority of questions are the same on each version of the questionnaire, which allows the two populations to be analyzed together for most issues, there are some important differences. Perhaps the most important difference has to do with the selection of the respondent. The parent or representative payee is the respondent for sample members under the age of 18 and for young adults aged 18 or older living at home. Young

**Table 1.**  
**Number of completed interviews in the NSCF, by analysis group**

Group	Unweighted	Weighted
Children and young adults receiving SSI in December 2000	5,006	1,053,376
On SSI in December 1996	3,271	572,415
Not on SSI in December 1996	1,735	480,961
Children and young adults receiving SSI in December 1996	5,033	932,544
Subject to redetermination under welfare reform	3,185	282,067
Continued	1,549	120,438
Denied	1,636	161,629
Not subject to redetermination under welfare reform	1,848	650,477
Children and young adults who formerly received SSI and denied applicants	1,767	1,862,579
Number in sample <sup>a</sup>	8,535	3,276,084

SOURCE: Authors' calculations from the National Survey of SSI Children and Families (NSCF), interviews conducted between July 2001 and June 2002.

NOTE: Tabulations do not include the 191 sample members who were incarcerated at the time of the NSCF interview. An abbreviated interview was conducted with the parent or guardian of incarcerated sample members.

- a. Sample members may appear in more than one group. The unweighted number of observations in the sample is composed of three groups: (1) the 1,735 children and young adults receiving SSI in December 2000 who were not receiving SSI in December 1996, (2) the 5,033 children and young adults receiving SSI in December 1996, and (3) the 1,767 children and young adults who formerly received SSI and denied applicants.



**Table 2.****Number of completed interviews in the NSCF, by analysis group and recipients' characteristics**

Characteristic	Children and young adults receiving SSI in December 2000				Children and young adults receiving SSI in December 1996			
	Unweighted		Weighted		Unweighted		Weighted	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
Total	5,006	100.0	1,053,376	100.0	5,033	100.0	932,544	100.0
Age in December 2000 <sup>a</sup>								
0–5	532	10.6	154,492	14.7	76	1.5	36,583	3.9
6–12	1,535	30.7	357,538	33.9	1,410	28.0	315,854	33.9
13–16	1,108	22.1	230,332	21.9	1,340	26.6	244,073	26.2
17–18	960	19.2	97,745	9.3	926	18.4	111,466	12.0
19 or older	871	17.4	213,269	20.2	1,281	25.5	224,569	24.1
Age in December 1996 of children and young adults receiving SSI in December 1996								
0–5	...	...	...	...	625	12.4	188,186	20.2
6–12	...	...	...	...	2,202	43.7	408,046	43.8
13–16	...	...	...	...	1,411	28.0	222,655	23.9
17–18	...	...	...	...	776	15.4	105,272	11.3
19 or older	...	...	...	...	20	0.4	8,385	0.9
Sex								
Male	3,089	61.7	639,275	60.7	3,265	64.9	590,397	63.3
Female	1,917	38.3	414,101	39.3	1,768	35.1	342,146	36.7
Main reported health condition								
Mental disorder								
Retardation	393	7.9	74,416	7.1	330	6.6	57,083	6.1
Behavioral	150	3.0	26,823	2.5	156	3.1	21,660	2.3
Other	1,953	39.0	375,138	35.6	2,026	40.3	315,706	33.9
Physical (nonmental) disorder	1,792	35.8	418,976	39.8	1,644	32.7	361,438	38.8
Other	427	8.5	99,460	9.4	360	7.2	77,379	8.3
None reported	22	0.4	4,330	0.4	35	0.7	7,204	0.8
Missing	269	5.4	54,233	5.1	482	9.6	92,074	9.9
Reported general health status								
Excellent	530	10.6	119,257	11.3	576	11.4	116,810	12.5
Very good	739	14.8	157,015	14.9	742	14.7	149,200	16.0
Good	1,713	34.2	365,154	34.7	1,700	33.8	316,281	33.9
Fair	1,498	29.9	308,410	29.3	1,467	29.1	261,357	28.0
Poor	506	10.1	99,543	9.4	532	10.6	85,608	9.2
Missing	20	0.4	3,997	0.4	16	0.3	3,287	0.4

SOURCE: Authors' calculations from the National Survey of SSI Children and Families (NSCF), interviews conducted between July 2001 and June 2002.

NOTES: Tabulations do not include the 191 sample members who were incarcerated at the time of the NSCF interview. An abbreviated interview was conducted with the parent or guardian of incarcerated sample members.

... = not applicable.

a. Age in December 2000 is between 6 and 18 months less than age at the time of the interview.

adults living in their own home are the respondent, but they can designate a proxy respondent if they are unable to complete the interview because of a disability. Another important difference between the children's questionnaire and the one for young adults is that the latter contains more detailed questions about preparations for the transition to adulthood. For example, both versions ask about involvement with special education programs and whether an individual education plan was prepared. The questionnaire for young adults also asks about involvement in vocational training programs and other programs geared toward providing youth-to-adult transition services for individuals with disabilities. It also gathers information about the young adult's employment, earnings, and other sources of income. (This information is also gathered for the parents if the young adult is living at home.) In contrast, the children's questionnaire focuses on the employment, earnings, and income sources of the parents.

The major sections of the questionnaire include disability status and functional limitations; health care utilization; health insurance; education and training; programs and services; impact on family, household members, and self; SSI experience; employment and earnings; work and child care; unearned income and assets; and housing and transportation. (For additional details about the content of each section of the questionnaire, see Box 1.)

The interviews took approximately 70 minutes on average to complete. Interviews with children who were living with both parents took slightly longer because employment and earnings data were collected for both parents. Interviews were significantly shorter with respondents who reported that the sample member was not disabled, did not receive SSI benefits, or made relatively little use of the health care system.

Although the questionnaire contains many questions that were designed to address topics specific to children and young adults with disabilities who receive SSI or topics of particular interest to SSA, many questions were drawn from existing surveys. For example, questions regarding employment, earnings, other income sources, and assets were taken from the Survey of Income and Program Participation. Questions regarding health status and functional limitations were drawn from the National Health Interview Survey. The purpose of using questions from these other sources is to preserve comparability between the NSCF and other major surveys. Nonetheless, comparisons between the NSCF and other surveys must be made with great caution because of different modes of interviewing and different sampling, weighting, and imputing techniques.

## ***Data Collection***

NSCF interviews were conducted between July 2001 and June 2002. Most interviews were conducted during the second half of 2001. A number of steps were taken to maximize the quality of the data collected in the NSCF. For example, the survey was designed for mixed-mode data collection to improve response rates and to avoid potential biases resulting from telephone-only surveys. Costly efforts to locate sample members were also undertaken to help achieve a high survey completion rate. These efforts are discussed briefly below. See Appendix A for a discussion of the randomized experiment that was conducted to test different response incentives to boost cooperation among interview subjects who were located.

### ***Mixed-Mode Data Collection***

The primary method for collecting data was computer-assisted telephone interviewing (CATI). Of the 8,726 interviews (8,535 completed interviews plus 191 abbreviated interviews with sample members who were incarcerated at the time of the NSCF interview), 7,285 were completed by CATI. For sample members who could not be reached by telephone (for example, a correct telephone number could not be found) or who could not complete the interview by telephone (for example, the respondent's disability prevented him or her from responding by telephone, language barriers), field interviews were attempted using computer-assisted personal interviewing (CAPI). The remaining 1,441 interviews (16.5 percent of the total) were conducted using CAPI. The same version of the questionnaire was administered for both CATI and CAPI interviews to minimize potential mode effects.

### ***Locating Sample Members***

The biggest challenge in conducting the survey was locating sample members. Although SSI administrative data were used to identify the telephone number and address of sample members, about 70 percent of the sample had addresses that were no longer valid. Even among SSI recipients at the time of the survey, nearly 50 percent had invalid telephone numbers or addresses. Numerous methods were used to locate sample members, including searches of commercially available databases and ground work by NSCF field interviewers in the sample member's last known neighborhood. In the end, about 84 percent of the NSCF sample was located for interviewing. Despite the overall success of these efforts, nearly three-quarters of final nonresponses in the NSCF were because the sample member could not be found.

**Box 1.****Content of the NSCF questionnaire**

Section	Description
Disability status and functional limitations	Screens children and young adults for the presence of a health condition, then follows up with questions about the nature, severity, and duration of the condition. The questions will allow the construction of disability indices by severity of reported limitations. Using several of the items together will allow classification of respondents into severity groups and facilitate comparisons with other national surveys.
Health care utilization	Collects descriptive information on the child's or young adult's frequency of use of doctors, hospitals, emergency room care, and prescription drugs. Asks questions about the family's out-of-pocket health care expenses and the child's or young adult's unmet health care needs.
Health insurance	Asks questions about the type of health insurance the child or young adult has, who pays for the coverage, and episodes when the child was without health coverage (if any).
Education and training	Collects data on the child's or young adult's educational attainment and receipt of special education, early intervention, and vocational education services. Particularly important for analyses of the youth-to-adult transition process.
Programs and services	Asks questions about programs and services used or needed by the families of SSI children and young adults, including therapy, respite care, and family counseling. Collects data about who pays for the services, unmet needs for services, and the out-of-pocket costs.
Impact on family, household members, and self	Asks questions about quality-of-life issues. Items cover food, housing, and financial security. Items are included on the child's behavior and social interactions and the impact of having a disabled child on the household's interactions and living arrangements.
SSI experience	Covers receipt of SSI benefits and the family's experience with redeterminations and the appeals process. Other items ask about how the household uses the SSI benefit. Also asks about the family's familiarity with and use of SSA-sponsored work incentive programs for SSI recipients (for example, Plan for Achieving Self-Support, Individual Development Account, earned-income exclusions).
Employment and earnings	Covers the employment of parents and SSI children and young adults who are employed. Questions ask about the type of work performed, type of employer, hours worked, and wages earned. Questions address the effect of having a disabled child on parental labor force participation, and the ability to work and work experience of young adults.
Work and child care	Asks about child care while parents are working or attending school. Questions address who provides the care, the number of hours of care provided each week, the need for specialized care, satisfaction with care, and the cost to the family.
Unearned income and assets	Asks detailed questions on the household's receipt of unearned income including government benefits (food stamps, Temporary Assistance for Needy Families, foster care payments, unemployment compensation) and other unearned income (child support, pension payments). Questions ask who in the household received the benefit or payment and the amount received last month. Other questions ask about the assets of the parent or guardian (and spouse or partner, if any) and overall debt burden.
Housing and transportation	Asks questions about the type of housing the child or young adult lives in, the cost of the housing, and the availability of or need for modifications to accommodate persons with disabilities. Also asks about types of transportation used and the child's or young adult's need for special modifications when using public transportation.

SOURCE: National Survey of SSI Children and Families (NSCF) questionnaire, available at <http://www.socialsecurity.gov/disabilityresearch/nscf.htm>.

NOTE: SSI = Supplemental Security Income; SSA = Social Security Administration.

### *Cooperation of Sample Members*

Once sample members had been located, obtaining their cooperation was very successful. Observations of a small number of interviews suggest that respondents were quite willing to answer the NSCF questions and that many seemed happy to “tell their story” to an interested listener. To encourage cooperation, letters sent in advance of the NSCF interview informed sample members that they would receive a response incentive of \$10 upon completion of the interview (see Appendix A for details). The cooperation rate is the number of completed interviews as a percentage of the number of sample members who were located. The weighted cooperation rate was 90.6 percent overall and 93.7 percent among SSI recipients at the time of the survey.

An important issue with respect to the quality of the survey data is the pattern of survey noncompletion. Random patterns of noncompletion produce no selection bias in survey estimates, although they lower efficiency because they reduce the sample size. However, if survey noncompletion is systematically and substantially associated with relevant measured or unmeasured characteristics, concern arises about whether the sample is representative. Appendix B provides three (weighted) measures relevant to assessing potential selectivity bias by a number of characteristics. Those measures are the percentage who were located, the percentage who cooperated among those who were located, and the overall completion rate, which is the product of the first two sources of sample selectivity.

Overall, 81.9 percent of the sample members were located. Among those who were located, 90.6 percent completed the interview. The product of these two factors is the overall completion rate; 74.4 percent is a slightly adjusted version of this product.<sup>4</sup> The data also show that there was substantially more variation in the proportion who were located than in the proportion who cooperated among those who were located. The rate located was relatively low for nonrecipients (those who, according to administrative records, had no payment) and those who had recently moved (78.2 percent and 77.2 percent, respectively) and relatively high for those with a grandparent as a representative payee (91.3 percent). Both sources of noncompletion contributed to the substantial overall difference (15 percentage points) in the completion rate between SSI recipients and sample members who were not receiving SSI at the time of the survey.<sup>5</sup> They also contributed to the relatively high completion rate among sample members for whom a mental impairment was listed as the reason they applied for SSI, but the differences were less marked. The subgroup differences for most variables were relatively modest. Note that these data reflect differences before

the nonresponse adjustment and weight adjustments that were used to address these issues, as discussed below.

### *Procedures for Imputations, Weights, and Variance Estimation*

---

Mathematica statisticians developed imputation procedures, weights, and variance estimation procedures. Imputed values allow the use of observations with item nonresponse in statistical and econometric analyses. Weights are necessary for unbiased estimation of population means and other statistics. The variance estimation procedures developed by Mathematica are important tools for deriving unbiased estimates of the precision of statistical estimates (sampling variance) in the presence of the complex NSCF sample design, which makes the assumption of simple random sampling in variance estimation inaccurate.

#### *Imputing Values for Missing Data Items*

The rates of item nonresponse and the methods used to impute missing values are important indicators of the quality of the survey data files. In the NSCF, item nonresponse rates were generally low, and sophisticated imputation procedures were used. Administering the questionnaire by using computer-assisted instruments substantially reduced the extent of item nonresponse. With computer-assisted instruments, interviewers are automatically routed through the questionnaire, questions cannot be left blank, and automated checks of consistency and reasonableness are implemented. Nevertheless, missing data are sometimes generated because questions are mistakenly not answered and respondents sometimes provide “don’t know” responses or refuse to answer a question. In addition, the value of derived variables is missing if any of the component variables were subject to item nonresponse.

Because of the length of the NSCF questionnaire and the number of variables on the data file, the cost of imputing values for missing data on all variables appeared to be prohibitive. Mathematica therefore focused its imputation efforts on variables for earned and unearned income, assets, and out-of-pocket expenditures. Not only are those variables of great importance for planned analyses with the NSCF data, but item nonresponse was relatively high for some of them. In addition, summary variables, such as total income and assets, were derived from a number of survey items, each one of them a potential source of a missing value for the summary measure. For these reasons the potential payoffs of high-quality imputations were especially large for data items on income, assets, and out-of-pocket expenditures. In contrast, variables related to the sample member’s type



of impairment and functional limitations, for example, are poor candidates for imputation because their correlation with other observed characteristics is inherently complex.

Altogether, 97 variables were imputed using one or more of the following methods: logical imputation, hot-deck imputation, and regression-based imputation. For each variable for which missing values were imputed, an imputation flag was created and added to the NSCF database. For each observation, this flag indicates whether or not the value for that observation was imputed and the imputation method that was used. Potter and Diaz-Tena (2003) provide additional details on the imputation methods. Analysts interested in using variables for out-of-pocket expenditures, health care utilization, employment, unearned income, and assets are encouraged to consult Potter and Diaz-Tena (2003, Table IV.2) to assess both the extent of item nonresponse and the resulting portion of missing values for derived variables. The NSCF data documentation also provides detailed information related to item nonresponse and imputations for other survey variables in the file.

### ***Weighting the Sample***

Mathematica developed analysis weights for the NSCF using a three-step process. First, they developed sampling weights, which are simply the inverse of the probability of selection for each sample member. The sampling weights thus account for both the selection of primary sampling units and the individual selection probabilities within each stratum. The initial sampling weights ranged from 29.7 to 510.7.

Second, they made two adjustments to the initial sampling weights for nonresponse: one was based on the ability to locate the sample member, and the other was based on the cooperation of sample members once they had been located. Weighted logistic regression models were estimated to develop a propensity score for locating the sample member and a propensity score for response among those located. Covariates used in the logistic regression models include the sample member's personal characteristics, current and previous SSI status, geographic region and urban/rural status, and the personal characteristics of an adult living with the sample member. Potter and Diaz-Tena (2003) provide detailed results from the logistic regression models. The nonresponse adjustment is the inverse of each propensity score, which was then multiplied by the initial sampling weight to obtain the response-adjusted weight.

Finally, Mathematica poststratified the response-adjusted weights to match the weighted sums across 11 analytic populations and by sex, with control totals for groups calculated from SSA administrative data. The weight variable included on the data file provides the final

analysis weight for each sample member, reflecting the adjustments described above.

### ***Variance Estimation***

The NSCF sample design uses a complex, multistage procedure for sample selection. As a result, the standard variance calculations assuming a simple random sample (SRS) will underestimate the true variance. Therefore, Mathematica developed two sets of variance estimation specifications for the NSCF.<sup>6</sup> Sample code for implementing each procedure is included in the NSCF User's Manual (Gillcrist and Edson 2004).

Means and standard errors calculated using the Taylor-series linearization for several NSCF variables of interest are shown in Table 3, which also presents the Taylor-series standard errors as a percentage of the SRS standard errors. The SRS standard errors are smaller than the standard errors that account for the complex NSCF sample design and thus would lead to overly optimistic conclusions about the precision of estimates. Accounting for the complex sample design clearly makes a substantial difference in most cases, and the magnitude of the difference varies greatly depending on the particular variable in question. If the bias involved in using the uncorrected SRS estimate of standard errors was roughly similar across the board, a single value of the design effect could be used as a rough way to correct for the complex survey design. Although this assumption may be correct in some surveys, the nature of the NSCF sample design would make it incorrect in this survey. For example, Table 3 shows that the SRS estimate of standard errors underestimates the true standard errors for health insurance coverage among children receiving SSI in December 2000 by a factor of 1.25; in contrast, the corresponding underestimate of the percentage who are black is much larger, 4.08. Because of the importance of accounting for the complex NSCF survey design, analysts should use appropriate software to calculate corrected standard errors. Potter and Diaz-Tena (2003) explain the application of two widely used techniques—the Taylor-series linearization procedure and the balanced repeated replication (BRR) procedure—for the NSCF data set.

### ***Data Products***

Mathematica prepared several major data-related products for use by NSCF researchers—a master data file, a User's Manual, a report on the quality of the NSCF data, a report on weighting and imputations, and a report comparing data from the NSCF with data on children receiving SSI from other national surveys. They also prepared a public-use version of the NSCF data set and documentation that is available to researchers

studying disabled children enrolled in the SSI program. The data set is available at <http://www.socialsecurity.gov/disabilityresearch/nscf.htm>.

SSA's master data file contains all of the information collected by the NSCF. It also contains administrative data items that were used to design and control the sample and numerous derived and recoded variables to make the data more user friendly. For example, responses to the open-ended questions on type of impairment were recoded following the International Classification of Diseases (ICD) coding scheme and the SSA diagnosis coding scheme. Administrative data from other SSA systems can be appended to the NSCF as

needed for individual SSA projects.

The User's Manual prepared by Mathematica (Gillcrisp and Edson 2004) is a critically important resource for successfully using the NSCF for research and analysis. In addition to the detailed codebook, the User's Manual includes a discussion of the NSCF sample design, the questionnaire design and content, and procedures for collecting data, constructing and editing variables, and estimating variances. It also provides instructions for subsetting the data to identify various subpopulations of potential interest to researchers. Used in conjunction with three other reports—on data quality (Gillcrisp, Kasprzyk, and Mitchell 2004); on weighting, nonresponse adjust-

**Table 3.**  
**Means and standard errors of selected variables for children and young adults receiving SSI in December 2000 and children and young adults receiving SSI in December 1996**

Variable	Mean	Standard errors corrected for complex sample design <sup>a</sup>	Corrected standard error as a percentage of uncorrected standard error <sup>b</sup>
Children and young adults receiving SSI in December 2000			
Black (percent)	43.9	2.9	408
Hispanic (percent)	15.7	1.8	350
Earnings of parent or guardian in previous month (dollars)	1,073.83	45.59	217
Total household income in previous month (dollars)	1,953.69	49.40	233
Total assets of parent or guardian and spouse or partner, if any (dollars)	2,531.68	270.08	179
Household receipt of food stamps last month (percent)	30.8	1.3	194
Health insurance coverage (percent)	97.6	0.3	125
Total out-of-pocket health care expenses in previous 12 months (dollars)	264.12	32.35	131
Children and young adults receiving SSI in December 1996			
Black (percent)	43.7	2.8	393
Hispanic (percent)	14.1	1.8	363
Earnings of parent or guardian in previous month (dollars)	1,185.31	46.28	193
Total household income in previous month (dollars)	1,961.25	45.38	187
Total assets of parent or guardian and spouse or partner, if any (dollars)	2,032.20	189.54	162
Household receipt of food stamps last month (percent)	29.8	1.4	218
Health insurance coverage (percent)	89.1	0.8	172
Total out-of-pocket health care expenses in previous 12 months (dollars)	291.04	32.73	124

SOURCE: Authors' calculations from the National Survey of SSI Children and Families (NSCF), interviews conducted between July 2001 and June 2002.

NOTE: Tabulations do not include the 191 sample members who were incarcerated at the time of the NSCF interview. An abbreviated interview was conducted with the parent or guardian of incarcerated sample members.

- a. Calculated using the Taylor-series linearization, which corrects for the complex sample design.
- b. Expressed as ratios, these percentages equal the square root of the statistical design effect. For example, the figure of 125 percent for health insurance coverage expressed as a rate (1.25) corresponds to a design effect of 1.56. Uncorrected standard errors (not shown) are calculated assuming a simple random sample (SRS).

ments, and imputation (Potter and Diaz-Tena 2003); and comparing estimates from the NSCF to estimates for children receiving SSI from other national surveys (Ireys and others 2004)—users will be prepared to study the wealth of information on children receiving SSI that is available in the NSCF.

### ***Appendix A. Experiment with Response Incentives***

Offering an incentive to participate in surveys like the National Survey of SSI Children and Families is standard practice. In most cases, a small payment by check is offered in advance to sample members and is paid once they have completed the survey. In addition to this standard approach, Mathematica proposed an experiment to test two alternative incentives—a point-of-sale (debit) card and a prepaid telephone card. About 70 percent of NSCF sample members were randomly selected to receive the standard check payment, 15 percent to receive a debit card, and the remaining 15 percent a prepaid telephone card. The experiment did not include a no-incentive control group, based partly on the presumption that the offer of an incentive payment cannot produce a negative effect on the willingness of sample members to cooperate. Sample members were sent a letter in advance telling them which type of payment they would receive and its value (\$10).<sup>7</sup> Mathematica tracked the completion rate, usage rate, administrative issues, and costs of each type of payment.

For two reasons, the key outcome of interest from the government's perspective is the effect on the final survey completion rate. First, increasing the survey completion rate is believed to improve the overall quality of the survey product, presumably by reducing the potential for selectivity bias that arises from patterns of survey noncompletion.<sup>8</sup> Second, response incentive payments, if successful, may be highly cost-effective relative to other methods of increasing survey completion rates. For example, shifting from a mixed mode of data collection to face-to-face interviews may be rather costly.

The final completion rate was statistically significantly higher for sample members who received a debit card (79 percent) than for those who received a telephone card (75 percent). The differences between the response rate for debit cards and the rate for checks (77 percent) and between the rates for checks and telephone cards were not statistically significant. Given this overall pattern and assuming that the offer of telephone cards did not reduce the willingness to cooperate relative to the hypothetical “no incentive payment” counterfactual, these results suggest that the use of debit cards was successful and should be regarded as the preferred method. Although the exact magnitude of this effect is highly

uncertain, based on this evidence the best point estimate is an increase of 4 percentage points attributable to the use of debit cards versus telephone cards. This difference is substantial, suggesting that using a debit card as an incentive is clearly cost-effective.

Our conclusions differ from those reached by Mathematica's Mitchell, Lamothe-Galette, and Potter. They conclude (2003, 4) that “In the end, POS [debit] and telephone cards did not perform as dependably or economically overall as checks.” The differences stem primarily from differences in the key outcomes of interest. Whereas the Mathematica authors focused on operational difficulties and the unit cost of the three interventions from the research firm's perspective, we focus on the key outcome of interest for the government—the net effect on survey completion rates.

In addition, there are three areas of disagreement between our interpretation and theirs. First, with respect to cost, Mitchell, Lamothe-Galette, and Potter (2003, 3) conclude that “costs loom large” on the basis of differences between their estimate of the total unit cost of debit cards (\$12.21) and checks (\$11.20). We argue that the \$1 difference in unit cost is minuscule in comparison with the potential government savings from using this low-cost method to boost survey completion rates as opposed to using more expensive methods (for example, shifting from telephone to field interviews). In addition, Mitchell and colleagues ignore cost differentials arising from differences in respondents' use of the three types of incentive payments. Once the probability of using the incentive payment is considered, the pattern of unit cost estimates dramatically shifts in favor of the debit cards.

Second, Mitchell, Lamothe-Galette, and Potter take the relatively low usage rate of debit cards (47 percent) versus checks (85 percent) as *prima facie* evidence of the ineffectiveness of using debit cards. This interpretation is problematic because these types of experiments involve a “deadweight loss” (that is, behavior may be unaffected by the response incentive). If a person would complete the interview whether he or she received an incentive payment or not, there is no gain in completion rate attributable to the incentive—the \$10 incentive payment is wasted. It is the marginal respondent—the one whose behavior is positively altered by the offer of the incentive—who counts. In this study, the deadweight loss may be as high as 75 percent. The experimental results suggest that the debit cards were more successful with marginal respondents.

Finally, our interpretation of the statistical evidence is different. In our interpretation the key result is that the completion rate using debit cards (79 percent) was statistically significantly higher than the rate with telephone cards (75 percent), which can be taken as an

upper-bound estimate of the “no response incentive” counterfactual. The fact that neither number is significantly different from the response rate using checks (77 percent) suggests that one cannot reject the hypothesis that the true effect of the debit cards may be only 2 percentage points. That would still be a substantial gain in completion rates given the low cost of any of the three interventions. Our interpretation is also supported by the fact that the completion rate using debit cards was 5 percentage points higher than the rate with telephone cards at the end of the first month of data collection—the time when behavior is most likely to have been affected by the experiment with incentives. By contrast, there is no difference between the rates for checks and telephone cards at the end of the first month. The completion rate among debit card recipients clearly exceeds that of check and telephone card recipients for months 1 through 11 of data collection, as reported by Mitchell, Lamothe-Galette, and Potter (2003, Table 1).

## Appendix B. NSCF Location and Completion Rates

Table B-1 provides weighted location and completion rates for NSCF sample members by characteristics of the sample members. The table was adapted from Table III.3 of Potter and Diaz-Tena (2003).

### Notes

*Acknowledgments:* The authors are grateful to Michael Compson, Edward DeMarco, Brian Greenberg, Susan Grad, Howard Iams, Paul Van de Water, and Martynas Ycas for helpful comments and suggestions. We would like to thank Paul O’Leary for his significant contributions during the design stage of the NSCF. Finally, we are grateful to the staff of Mathematica Policy Research, Inc., who worked so diligently to design and implement the NSCF, and to Susan Mitchell and Frank Potter for their leadership on this project.

**Table B-1.**  
**Weighted location and completion rates for NSCF sample members, by characteristics of sample members**

Characteristic	Weighted percentage located	Weighted percentage completed/located	Weighted percentage completed and ineligible
All sample members	81.9	90.6	74.4
Age			
0–17 years	82.7	91.3	75.7
18 years or older	79.9	89.0	71.4
Type of disability			
Mental	85.8	93.5	80.3
Physical	80.1	89.3	71.7
Race and ethnicity			
Black	80.9	91.0	73.8
White	81.7	90.4	74.1
Hispanic	79.5	88.6	70.6
Unknown	83.7	90.8	76.2
Sex			
Male	81.8	90.5	74.2
Female	81.9	90.7	74.6
Region			
Northeast	85.1	89.1	75.9
South	85.4	92.6	79.3
Midwest	82.8	93.4	77.3
West	84.3	92.4	78.2
Unknown	76.3	87.5	67.0
Urban			
Yes	81.5	89.9	73.4
No	83.1	92.7	77.2
Recently moved			
Yes	77.2	90.3	70.0
No	82.2	90.6	74.7

(Continued)



<sup>1</sup> Another goal of the NSCF was to provide data to support further evaluation of the effects of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (P.L. 104-193, otherwise known as welfare reform) on children receiving SSI. Although this is a secondary objective of the NSCF, it was an important consideration in defining the major analysis groups for the survey.

<sup>2</sup> The NSCF used a two-stage sample design based on a list frame derived from the Social Security Administration's administrative records. The final sample size was 11,971 cases. Potential respondents for 10,025 sample members were located and 9,242 respondents cooperated. Of those who cooperated, 516 were excluded because they were deemed ineligible for the survey (that is, they were deceased, living in a Medicaid institution, wards of the state, or living in Alaska, Hawaii, or the Northern Mariana Islands). This yielded a total of 8,726 interviews, 191 of which were for incarcerated sample members. Only an abbreviated interview was conducted with the parent or guardian of incarcerated sample members. Because of the limited nature of this information and the survey's focus on the noninstitutionalized population, the data presented in this article exclude incarcerated sample members, yielding a total sample of 8,535 completed survey interviews.

<sup>3</sup> The unweighted completion rate calculated by Mathematica was 77.2 percent. Using a different approach to calculate completion rates that removes ineligible sample members from both the numerator and denominator yields an unweighted completion rate of 76.2 percent. This second approach is advocated by the Office of Management and Budget.

<sup>4</sup> The product of the first two factors gives an overall completion rate of 74.2 percent. The reason for the difference of 0.2 percentage points is that the summary measure, following standard survey practice, adjusts the statistics by accounting for the number who were attempted to be interviewed but were found to be ineligible for the survey. The intuitive explanation of this practice is that in the case of ineligible sample members, neither the failure to locate respondents nor the failure of respondents to cooperate (conditional on having been located) contributed to the lack of a completed interview.

<sup>5</sup> The difference of 15 percentage points was derived on the basis of the variable for federal payment amount. Appendix B contains the relevant value for nonrecipients (69.6 percent). We calculated the corresponding figure for recipients as a weighted average of the percentage for those receiving less than \$500 in monthly payments (85.4 percent) and the percent-

**Table B-1.**  
**Continued**

Characteristic	Weighted percentage located	Weighted percentage completed/located	Weighted percentage completed and ineligible
Currently enrolled in school			
Yes	89.7	94.4	84.8
No	78.3	88.8	69.7
SSI status in December 1996			
SSI recipient and subject to redetermination under welfare reform			
Payments continued	84.3	92.9	78.4
Payments ceased	80.6	92.5	74.7
SSI recipient and not subject to redetermination under welfare reform	83.6	91.9	76.9
Not an SSI recipient	80.8	89.4	72.5
Federal payment amount (monthly)			
No payment	78.2	88.9	69.6
Less than \$500	91.1	93.7	85.4
\$500 or more	88.6	94.4	83.7
Years of SSI receipt			
0-4	79.9	89.9	72.1
5-9	85.7	92.2	79.1
10 or more	87.8	92.0	80.8
Type of representative payee			
Agency	82.2	96.4	79.6
Father	86.1	89.2	76.9
Grandparent	91.3	96.0	87.6
Mother	84.2	91.0	76.8
Other	78.1	89.3	70.0
Other relative	82.4	92.0	75.9

SOURCE: Adapted from Table III.3 of Potter and Diaz-Tena (2003).

age for those receiving \$500 or more in payments (83.7 percent). We used the count of total attempted interviews for the two groups presented in the first column of Table III.3 in Potter and Diaz-Tena (2003) as the weighting variable. This procedure yielded an overall completion rate of 84.3 percent for recipients. The difference of 14.7 percentage points between the two groups was rounded to 15 percentage points.

<sup>6</sup> The two procedures for variance estimation are the Taylor-series linearization technique and the balanced repeated replication (BRR) procedure. For technical details, see Levy and Lemeshow (1999). In this article, we use only the Taylor-series linearization and focus on highlighting the importance of correctly calculating standard errors in surveys with a complex sample design.

<sup>7</sup> One interesting possibility—offering a choice concerning the form of incentive payment—was not included in the experiment. The experiment was limited to testing the effects of a single payment level of \$10.

<sup>8</sup> Note that the magnitude of nonresponse error for each variable in a survey data file critically depends on the relationship between the patterns of nonresponse and the distribution of the given variable. In the extreme case of respondents forming a random subsample of the sample frame, nonresponse would reduce the statistical efficiency of the estimate but would not produce systematic error. However, if the difference between the distribution of a given variable for respondents and the (unobserved) distribution of that variable for nonrespondents is large, even a sample with a relatively high response rate could produce overall statistics that are substantially affected by nonresponse error. In general, higher response rates are thought to be helpful in reducing nonresponse error and increasing statistical efficiency. Nevertheless, increasing the response rate may make matters worse rather than better in some cases because of a strong association between variables affecting survey participation and the survey variables of interest. Analysts are encouraged to consider the potential magnitude of nonresponse error on an item-by-item basis.

## References

- Gillcrisp, Jennifer, and David Edson. 2004. *National Survey of SSI Children and Families User's Manual for Restricted and Public Use Files*. Princeton, NJ: Mathematica Policy Research.
- Gillcrisp, Jennifer, Daniel Kasprzyk, and Susan Mitchell. 2004. *Report on Data Quality in the National Survey of SSI Children and Families*. Princeton, NJ: Mathematica Policy Research.
- Ireys, Henry, Daniel Kasprzyk, Ama Takyi, and Jennifer Gillcrisp. 2004. *Estimating the Size and Characteristics of the SSI Child Population: A Comparison Between the NSCF and Three National Surveys*. Princeton, NJ: Mathematica Policy Research.
- Lazear, Katherine J., and Janice Worthington. 2002. *Supplemental Security Income (SSI) Family Impact Study*. Research and Training Center for Children's Mental Health, Florida Mental Health Institute, University of South Florida. Tampa, FL: University of South Florida.
- Levy, Paul S., and Stanley Lemeshow. 1999. *Sampling of Populations: Methods and Applications*, 3rd ed. New York: John Wiley & Sons.
- Mitchell, Susan, Colette Lamothe-Galette, and Frank Potter. 2003. *Survey Response Incentives for a Low-Income Population: What Works?* Mathematica Policy Research Issue Brief. Princeton, NJ: Mathematica Policy Research.
- Potter, Frank. 2001. *Report on Revised Sample Design: National Survey of SSI Children and Families*. Princeton, NJ: Mathematica Policy Research.
- Potter, Frank, and Nuria Diaz-Tena. 2003. *Weighting, Nonresponse Adjustments, and Imputation: National Survey of SSI Children and Families*. Princeton, NJ: Mathematica Policy Research.
- Rogowski, Jeannette, Lynn Karoly, Jacob Klerman, Moira Inkelas, Melissa Rowe, and Randall Hirscher. 2002. *Final Report for Policy Evaluation of the Effect of the 1996 Welfare Reform Legislation on SSI Benefits for Disabled Children*. DRU-2559-SSA. Report prepared for the Social Security Administration. Santa Monica, CA: RAND.
- Rupp, Kalman, Paul S. Davies, Chad Newcomb, Howard Iams, Carrie Becker, Shanti Mulpuru, Stephen Ressler, Kathleen Romig, and Baylor Miller. 2005/2006. "A Profile of Children with Disabilities Receiving SSI: Highlights from the National Survey of SSI Children and Families." *Social Security Bulletin* (66)2: 21–48.