National Center for Health Statistics
Division of Health Interview Statistics
State and Local Area Integrated Telephone Survey

Child Well-Being and Welfare Module Pilot Test
Minnesota, 1999

I. Introduction and Background

In 1994, the National Center for Health Statistics and the National Immunization Program, Centers for Disease Control and Prevention, implemented the National Immunization Survey (NIS) as one element of the Childhood Immunization Initiative. The State and Local Area Integrated Telephone Survey (SLAITS) was created as an expansion of the existing NIS to establish a broad-based ongoing surveillance system at the State and local levels. The expanded survey was designed in response to the critical need for a single standardized mechanism to provide health- and welfare-related population-based data at the State and local levels to track and monitor the health and well-being of children and adults.

Two SLAITS modules have been pilot-tested as of mid-1999. One module was concerned with family health, including issues of access to care, health status, and insurance. It was pilot-tested in two States in 1997.

The other SLAITS module concerns child well-being and public assistance program participation. It was pilot-tested in two states (Texas and Minnesota) in 1998-1999. This report details the methodology of the pilot test of the Child Well-Being and Welfare (CWBW) Module in Minnesota. Children currently receiving benefits through the federal Medicaid program (called Medical Assistance in Minnesota) or through MinnesotaCare (a state-subsidized health insurance program for low-income, working residents who don't have access to affordable coverage) were the population of interest.

National Immunization Survey

As of the early 1990s, preschool children in the U.S. were not adequately vaccinated for preventable diseases according to a variety of measures, although state laws requiring vaccinations for school entry had been credited with high vaccination rates in the school-age population. To address the problem, through the National Immunization Program, 78 states and local areas, known as Immunization Action Plan (IAP) Areas, were awarded grants for the improvement of vaccination levels of children by their second birthday. To monitor the use of these funds and vaccination levels of very young children within the IAP Areas and across the United States, the National Immunization Survey (NIS) was established.

Integration of SLAITS with NIS

From an administrative standpoint, the NIS presents an extremely challenging task, because of the large number of households the survey must screen in order to find its relatively rare target population of households with children between 19 and 35 months of age: nearly one million households are screened each year to obtain a statistically significant sample. At the same time, the NIS's large sampling frame of telephone numbers provides a cost-effective opportunity to survey other populations in addition to the rare population that eventually screens into the NIS itself. This program of research using the NIS sampling frame is called the State and Local Area Integrated Telephone Survey (SLAITS).

Goals of the Pilot Test

The prevalence of uninsured children is a critical public health issue that has attracted the attention of legislators and other public policymakers. Yet, the quality of health insurance reporting (especially in telephone surveys) has been the subject of much debate. The primary goal of this pilot test was to evaluate health insurance data collected by SLAITS. Toward this end, the CWBW Module was administered to parents or guardians of a sample of children known to be enrolled in public health insurance programs in Minnesota. By comparing SLAITS interview data with these enrollment records, it is possible to evaluate the accuracy of public health insurance reporting in SLAITS.

A second goal of the SLAITS pilot test was to describe the well-being of children enrolled in these health insurance programs in Minnesota. Therefore, a representative sample was selected randomly from a list of all children enrolled in these two programs. No NIS sample was used for this pilot test.

Key Features of the Pilot Test

Important features of the Minnesota pilot test included the following:

- Questionnaire to collect data on a maximum of two children, including health insurance coverage, child care arrangements, education, child well-being measures, welfare program participation, employment and earnings, demographic information, education, and income;
- 25-minute CATI interview for households with one child and 30-minute interview for households with two or more children;
- Respondent was the parent or guardian who knew the most about the sampled child's health care, child care, and education;
- 948 completed interviews (179 in households with one child and 769 in households with more than one child);
- Households requiring the administration of the screener and/or interview in Spanish included through the use of a Spanish translation of the questionnaire administered by bilingual interviewers;

II. Sample Design

The target population for this study was all children under age 18 in Minnesota currently enrolled in Medical Assistance or MinnesotaCare. On May 25, 1999 (i.e., two weeks prior to the scheduled start of data collection), the Minnesota Department of Human Services provided NCHS with a list of enrollees under age 18, including their birth dates, Social Security numbers (SSN), sex, counties of residence, and ZIP codes. Because data collection was scheduled to last 12 weeks, there was some concern that enrollees who were just shy of their 18th birthday would turn 18 (and thus be out-of-scope for the survey) before their households were contacted. Therefore, children whose 18th birthday would occur within eight weeks of the list construction were removed from the list. In addition, because the list was created prior to the start of data collection, very recently enrolled children were also effectively excluded from the population.

A sample of 1,836 children was randomly selected from this list of enrollees. This sample was necessarily larger than the survey sample desired because it was anticipated that some children on the list would have missing or inaccurate telephone numbers. Of the list of enrollees, 7.1% of the cases sampled did not have a telephone number. When initial calls were made, an additional 20.4% of the cases were identified as having an inaccurate telephone number (nonworking, nonresidential, fax/modem, temporarily not in service, or no children in the household contacted). When telephone numbers were missing or inaccurate, new telephone numbers were sought through the use of directory assistance, the National Change of Address Service, CD-ROMs of addresses and telephone numbers, and Internet searches.

In order to ensure that the health insurance section would be asked in an identical manner for these cases as for an RDD sample, the telephone numbers were called by interviewers who were blind to the enrollment status of this population. Once an adult aged 18 or older residing in the household was contacted, and after an explanation of the survey that included the authorizing legislation, confidentiality, voluntary and burden statements, a roster of children under age 18 in the household was obtained. At this point, the computer selected the child whose first name and birth date matched the name and birth date selected from the Medical Assistance / MinnesotaCare enrollment list. When birth date and name did not match, an algorithm selected the child that best matched on month of birth, year of birth, and first initial of first name. If no match was possible, but the enrollment list indicated a child enrolled at that telephone number, two children were randomly selected for interview. Full names and SSNs were then obtained for these children at the end of the interview. If a match still was not made, the household interview was omitted from the data file (N = 172 households). Interviewers and respondents were blind to this selection process. In fact, to ensure that interviewers were not aware of the special nature of this sample, NIS interviews were completed when NIS-eligible children were identified in the household. When they occurred, these interviews preceded the SLAITS interview. However, the data file does not include data collected from the NIS portion of this interview.

III. Questionnaire

The SLAITS Child Well-Being/Welfare Module questionnaire includes questions from the National Health Interview Survey (conducted by NCHS), the Survey of Income and Program Participation (conducted by the Census Bureau), the National Household Education Survey (conducted by NCES), the Survey of Income and Program Dynamics (conducted by Census), and the National Survey of America's Families (conducted by the Urban Institute). The screener integrated NIS screening questions and the CWBW screening questions: The respondent was asked the number of household members, the number under age 18 (if the number of household members was 1, this question was skipped), and the NIS eligibility question.

I.   *Identification of Child and Respondent*: Names and birthdates of household members under age 18 were rostered. Gender and race/ethnicity data were also collected.

II.   *Parents and Household Composition*: Roster of household members, identification of relevant adult(s) for the sampled child(ren), spouse or partner, relationship to sampled child, marital status, and race/ethnicity of relevant adult.

III. *Health Insurance*: Type of health insurance coverage, if any, and any interruption in health insurance during the past 12 months.

IV.   *Child Care Arrangements*: Type of child care arrangement, impact on employment, difficulties arranging child care, and cost of child care.

V.   *Education*: For sampled child and relevant adults, current and completed education; highest grade or year of school, degrees or certificates earned.

VI.   *Child Well-Being:* School experience (number of schools attended, school contact regarding behavioral or academic problems, repeated grades), activities (i.e., outings, sports teams, clubs or organizations), neighborhood/community safety, and parental stress.

VII. *Welfare Program Participation*: Participation in welfare programs during the past year, reasons for participation, work requirement for participation, length of participation, and amount of monthly payment.

VIII. *Employment & Earnings*: Employment during the past year for the child and the relevant adults (number of jobs, length of employment, hours per week, days per week, commuting time, and employment earnings), and reasons relevant adults not employed.

IX.   *Income*: Total family income, income from all jobs, and income from other sources.

X.   *Household Information:* Birth outside of the United States; U.S. citizenship; additional telephone lines; interruption in telephone service; the child's Social Security number, name, and date of birth.

CATI Programming

The questionnaire was programmed as a module of the NIS CATI questionnaire, making full use of the CATI system's ability to check whether a response is within a legitimate range, follow skip patterns, and employ pick lists to present response categories. The question series on multiple telephone lines was identical for the NIS and SLAITS portions of the interview. These questions remained in the NIS portion for the NIS-eligible households, but were included at the end of the SLAITS interview for NIS-ineligible households. For respondents who had an NIS-eligible child, the NIS income questions were asked near the end of the SLAITS portion of the interview.

Spanish Version

The questionnaire was translated into Spanish by one translator. The Spanish version was then translated back into English by a different translator. Discrepancies were resolved in consultation with the two translators. The Spanish version was then incorporated in the CATI questionnaire.

IV. Data Collection

Interviewer Training

Interviewers who had conducted the SLAITS Child Well-Being/Welfare interviews in Texas, from October 1998–March 1999, continued with the Minnesota data collection. To compensate for attrition, six additional interviewers were trained prior to start of Minnesota data collection.

Interviewing

An advance letter, printed in English on one side and Spanish on the other, was mailed to the sampled households. Data collection began on June 6, 1999 and was completed on August 23, 1999. A total of 948 household interviews, about 1,663 children, were completed. A small pretest was conducted on the first day of data collection to test the CATI system. Since no problems were encountered, these interviews (and any subsequent interviews from this small sample replicate) were included in the final data file. The number of calls made to complete an interview ranged from 1 to 77, with a mean of 8.3 calls and a median of 5 calls.

Interview Length

The 948 interviews averaged 33.4 minutes in length, with a mean of 27.5 minutes in the 179 households with one child compared with 34.8 minutes for the 769 households with two or more children.

Other Languages

It was determined that administration of the screener and/or interview in Spanish was required in 38 households. Interviews were successfully completed by a Spanish-speaking interviewer using the Spanish version of the questionnaire in 33 of these households. In 106 households, a language

other than English or Spanish was spoken (at least 63 of these being Hmong-speaking households). The interview was not attempted in these households.

Breakoffs and Refusals

Interviews were completed in 213 households that had initially refused to participate (22.5% of the completed interviews). Of the 242 cases that were finalized as refusals, 35% broke off at the introduction, another 19% at the question regarding whether the respondent lives in the household and is over 17 years old, and 12% at the question regarding the number of people who live in the household.

Response Rates

Among cases where a household was determined to include a child, the interview completion rate was 86.1%. Application of the CASRO response rate calculation used for RDD telephone surveys to this sample resulted in response rates of 65.5% and 67.1%.

In both calculations, the classification of the households where a language other than English or Spanish would have been required to complete the interview has an impact on the calculation of the final response rate. In the calculations above, these cases are treated as unresolved cases. If, on the other hand, these cases are classified as out-of-scope cases, the response rates as calculated for an RDD sample are 70.6% and 72.0%.

Matched Medicaid Children

Among the 948 sampled children, 776 (81.9%) had a match with the Medical Assistance / MinnesotaCare enrollment list. (The public release data set includes only the data for these 776 children.) Looking only at the 504 matched cases identified on the enrollment list as Medical Assistance participants, 415 (82.3%) were reported to be insured by Medicaid or Medical Assistance during the interview.

V. Data Files

A SAS file containing the household-, family-, and child-level data variables was created. This file contains one record per sample child, with all information about the child's household on that record.

Editing

Concurrent with the development of the CATI questionnaire for the Texas CWBW Module data collection phase, a detailed plan for checking and editing the data in the CATI instrument was developed. The intention was to design into the CATI software consistency checks across data elements, valid range codes, and a method to identify incorrect codes entered by interviewers. To the extent that the CATI software could be developed to perform these tasks, the efficiency of post-survey data cleaning and processing was increased.

The CATI system was designed to perform a number of edits as an interviewer enters data into the computer system. These edits dealt with errors that could be reconciled while the respondent was on the telephone and focused, in particular, on items critical to the conduct of the study. The CATI edit specifications were designed to correct respondent error during the interview (for example, a respondent saying five children lived in the household, but only listing four names on the roster) and to identify and correct data-entry error by interviewers (for example, a 15-year old child is reported as coming to live in the USA in 1989, but the interviewer attempts to enter 1979, a year prior to the respondent's birth). To the extent possible without making the CATI system overly complicated, out-of-range and inconsistent responses resulted in a warning screen for the benefit of the interviewer, who was trained to correct errors as they occurred. These messages were designed primarily to prevent interviewer errors such as data entry errors and respondent errors, not to challenge respondents who gave logically inconsistent responses.

The two main types of CATI edits were range checks and consistency checks. A range violation would result in visual notification to the CATI interviewer (a pop-up box). In most cases the interviewer would have to enter a valid response in order to continue the interview. However, some out-of-range responses would produce a warning, and the interviewer would be instructed to verify the answer provided by the respondent. If the respondent confirmed the out-of-range value, the interviewer was allowed to continue. A consistency violation would result in visual notification to the interviewer (a pop-up box), indicating that an inconsistency between two responses had been detected. The interviewer would then have the opportunity to change one or both of the values entered. In some cases the interviewer had the option to proceed if the respondent confirmed the inconsistent values.

There are trade-offs between, on the one hand, incorporating every possible type of error check into a CATI system and, on the other hand, overall performance of the CATI system and the use of development resources. To reconcile this trade-off, post-CATI edits were developed to resolve problems that did not require access to the respondent.

After the pre-programmed edits were run, frequency distributions of all the variables in each of the files were produced and reviewed. The range of permissible values for each variable were examined for any additional invalid values or unusual distributions. Invalid values, where they occurred, were blanked out. If blank values already existed for a variable, they were checked to see whether they were allowable, due to legitimate skips, or occurred in excessive numbers. When blank values were the result of skip patterns, a "legitimate skip" code was used. Other records that were missing responses for unknown reasons were left missing. Any unanticipated consistency problems that were identified during the post-CATI editing were left inconsistent because these logic problems could not be resolved without further access to the respondent.

Edits to Protect Confidentiality

The Public Health Service Act (Section 308d) provides that data collected by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), may be used only for the purpose of health statistical reporting and analysis. Any effort to determine the identity of any reported case is prohibited by this law. NCHS does all it can to assure that the identity of data subjects cannot be disclosed. The risk of inadvertent disclosure of confidential information about individual respondents is higher with a publicly released data set having both

detailed geography variables and a detailed and extensive set of survey observations. Coarsening a data set by dropping survey variables, collapsing multiple variables into one, collapsing response categories for other variables, and/or introduction of noise in the data are common techniques to reduce the risk of inadvertent disclosure.

In the SLAITS data set, no geographic information below the state level has been reported. In addition, to reduce the possibility of identification of a family with an unusual structure, most of the data on the relationships between the focal child and all other family members has been removed. The only relationship information remaining concerns the relationship between the focal child and the adults in the household whose income or education may have a direct benefit for this child. We call these people "relevant" adults. Adults were categorized as relevant if they were (a) the child's mother or father, (b) partners of the child's mother or father, (c) the guardian who is the respondent (if not the mother or father), and (d) partners of the guardian who is the respondent. The maximum number of relevant adults in any one household is three.

To protect respondents' confidentiality, the following data were also edited:
• The child's age is reported in categories encompassing 2 or 3 years.
• The biological mother's age has been bottom-coded at 18 years or less and top-coded at 45 years or more.
• The response categories for race and ethnicity variables have been reduced to just two: non-Hispanic white and other.
• Education level for adults has been collapsed to four categories: up to 3 years high school, 4 years high school or high school graduate, 1-3 years college, and 4 years college or more.
• Education level for children has been recoded to four categories: nursery/preschool/kindergarten, elementary, grades 6-8, and grades 9-12.
• The number of schools attended has been top-coded at five schools or more.
• The number of months worked in the past year has been collapsed into three categories: less than 6, 6-11, and 12.
• The number of current jobs has been top-coded at three jobs or more.
• The number of hours worked per week has been bottom-coded at less than 20 hours, top-coded at 70 hours or more, and reported within a 10 hour range otherwise.
• The number of minutes for a commute has been reported in 10 minute ranges.
• Amount spent on child care has been reported in $25 ranges.
• Family income has been reported as 4 categories: less than $15000, $15000 - $24999, $25000 - $34999, and $35000 or more.
• Individual income from employment has been suppressed.

In addition, risk of inadvertent disclosure of confidential information is higher if a publicly released data set has program participation data for which participant enrollment data files may also be available. To reduce the risk of disclosure, detailed information about TANF and food stamp enrollment was collapsed into broad response categories. Due to limited enrollment, all information about General Assistance, General Relief, and Section 8 rental assistance was suppressed.

Releasing data about Medicare and Medicaid eligibility also provides a risk of disclosure for those non-traditional enrollees (e.g., children in Medicare) who may be receiving benefits (e.g., due to a disability). As an extra precaution to prevent positive identification of any individual enrollees, health insurance variables were collapsed into three categories: public coverage (including Medicaid, Medicare, Medi-Gap, Indian Health Service, and military), private coverage (including coverage obtained through employment or unions or purchased directly), and other coverage (including single-service plans, MinnesotaCare, General Assistance Medical Care, and any other kind of health insurance or health care plans).

VI. Guidelines

With the goal of mutual benefit, NCHS requests that recipients of data files cooperate in certain actions related to their use.

Any published material derived from the data should acknowledge NCHS as the original source. The suggested citation, "Source: National Center for Health Statistics, State and Local Area Integrated Telephone Survey, Child Well-Being and Welfare Module, Minnesota, 1999," should appear at the bottom of all tables. It should also include a disclaimer that credits any analyses, interpretations, or conclusions reached by the author (recipient of the file) and not to NCHS, which is responsible only for the initial data. Consumers who wish to publish a technical description of the data should make a reasonable effort to ensure that the description is not inconsistent with that published by NCHS.

As noted previously, the Public Health Service Act (Section 308d) provides that data collected by NCHS may be used only for the purpose of health statistical reporting and analysis. Any effort to determine the identity of any reported case is prohibited by this law. NCHS does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification are omitted from the data set. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users must:

1) Use the data in this data set for statistical reporting and analysis only.

2) Make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery.

3) Not link this data set with individually identifiable data from any other NCHS or non-NCHS data sets.

Use of the data set signifies users' agreement to comply with the above stated statutorily-based requirements.