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
Texas, 1998-1999

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 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. Children whose household incomes are below 200% of poverty is the population of interest. This pilot test of the Child Well-Being and Welfare (CWBW) Module is the subject of this report.

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 to monitor very young children's vaccination levels 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. At the same time, the NIS's large initial sample 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 initial sample is called the State and Local Area Integrated Telephone Survey (SLAITS).

Key Features of the Pilot Test

Important features of the Texas pilot test included the following:

- Use of Q4/1998 NIS sampling frame;
- Oversampling of low-income households with children (if necessary) to ensure that at least half the final sample consists of children living in households with income below 200% of the Federal Poverty Level;
- 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;
- 1,248 completed interviews (660 in households with one child and 588 in households with more than one child);
- Pretest of 30 respondents using the Computer Assisted Telephone Interview (CATI) questionnaire;
- NIS screener/interview completed prior to SLAITS portion of the interview
- Interviews of NIS-eligible households as well as NIS-ineligible households to be completed;
- 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;
- Survey estimates include a weight adjustment strategy to allow for the representation of households without telephones.

II. Sample Design

The study collected information on health and well-being of children under 18 years, and information on welfare program participation of households containing these children. The target population for this survey is all households with children under 18 years in Texas, and all children under 18 years in these households. The sample of households in scope for the survey was selected by screening telephone households via random-digit dialing (RDD). It is recognized that estimates based on an RDD survey alone may be biased due to noncoverage of nontelephone households. A method of adjustment to reduce this bias is described later in this report.

One of the requirements of the survey was to have a sample of 660 households with children under 18 years whose screened household income was below 200% of poverty level. It was also planned to have a sample of 588 households with children below 18 years and income above

200% of poverty. (These sample sizes reflect completed household interviews.) Children in each household were stratified into two age groups. All children from 0 to 5 years were in the first stratum and children between 6 and 17 years were in the second stratum. If children were present in both age groups in a selected household, then one child was selected at random from each age group. If there were children in only one age group, no more than two children were selected at random from that household.

Based on the distribution of households by number of children in Texas from the 1997 Current Population Survey (CPS), we expected to get a sample of 984 children from households with income below 200% of poverty level. This number was adjusted for possible misclassification of households as above or below 200% of poverty at the screening stage. This means that children classified below poverty in the sample could have come from households correctly screened as below poverty, and from households originally screened as below poverty but later recognized as above poverty. Previous studies indicated that 13% of the households screened as below 200% of poverty were above 200% of poverty based on detailed income data collected later. Similarly, 4.2% of the households originally classified as above 200% of poverty turned out to be below this poverty level based on detailed income data. Hence more low-income households were selected for the sample at the screening stage.

Screening Sample Size

A sample of 984 children from each group would give estimates with reasonable precision. For example, population percentages can be estimated within plus or minus 3.6 percentage points at a 95% level of confidence. Also, with this sample size, we can detect a difference of 6 percentage points with 80% power at a 5% level of significance, assuming that the base percentage is around 30%. (For this computation, we assumed a design effect of 1.3 due to clustering of children within households.)

The screening sample size is driven by the required sample size for the smallest domain. In this case, based on the 1997 CPS data, the smallest subpopulation was households with children below 18 and income below 200% of poverty level. We assumed an 85% response rate to the screening attempt and an 85% response rate to the data collection attempt. Also, 15.9% of the universe households were believed to be below 200% of poverty level. On the basis of these numbers, it was estimated that we needed to screen 5,742 households.

Subsampling of Households with an Income Above 200% of Poverty

This screening sample was expected to give more than the required sample size for households with children and income above 200% of poverty level. Therefore, it was decided to subsample the households with income above 200% of poverty level at the rate of 1 in 1.8. Later, during the actual sample selection, it turned out that we had overestimated the expected number of above-poverty households, and the subsampling fraction was set to 1.0. to achieve the desired sample size.

Sample Allocation and Selection

The sample of households selected for the Texas CWBW Module was essentially a subsample of the sample of households selected for the National Immunization Survey (NIS) in each IAP area in Texas. Therefore, the sampling frame for SLAITS in Texas was the same as the one used for the NIS, a brief description of which is given below.

The NIS employs a list-assisted RDD sampling frame. The list-assisted method uses the AT&T master tape of prefix area combinations of area codes and central office codes as the basis for constructing a sampling frame of banks of 100 consecutive telephone numbers. Following the creation of these 100-number banks, the most recent release of the Donnelly Marketing Information Services data file of residential, directory-listed telephone numbers is used to identify the banks of 100 numbers that have a very low probability of containing working residential numbers. These banks are deleted from the sampling frame as nonworking banks. Numbers in the remaining banks are matched with computer files of directory-listed residential and business lines to remove as many business lines as possible prior to telephone screening. The working banks in the sampling frame are updated quarterly to keep them accurate and current for the selection of the NIS sample each quarter. The list-assisted method yields a working residential number rate of 55% to 57%. A minor disadvantage of the list-assisted approach is that the removed banks of 100 numbers with zero directory-listed numbers may actually contain some working residential numbers. However, the proportion of residential telephone numbers excluded by using the list-assisted method is expected to be very small (Brick, et. al, 1995).[1]

The selection of telephone numbers for the Texas CWBW Module was achieved by selecting a certain number of replicates from the available replicates for the NIS in each IAP area in Texas. The number of replicates was determined by first calculating the number of sample households to be selected in each IAP area. Then, the number of telephone numbers to be called in each IAP area was determined based on the residential working number rate in that area.

There are 5 IAP areas in Texas. Assuming that the percentage of households with children and household income below 200% of poverty is the same in all IAP areas, we allocated the total number of households to be screened in Texas to each IAP area in proportion to the total number of households in that IAP area. The number of telephone numbers required to obtain the necessary number of households in one IAP exceeded that available in the NIS. Therefore, the number of households to be screened in that IAP area was set equal to the amount that could be obtained by calling the numbers available. The difference between the total number to be screened and the number assigned to that IAP was allocated in proportion to the number of households in the remaining IAP areas.

All households that were screened as below 200% of poverty were included in the sample. As indicated earlier, households screened as above 200% of poverty were subsampled in the beginning at the rate of 1 in 1.8, and later all households were included in the sample.

Selection of Children

---

1 Brick, J. Michael, Waksberg, Joseph, Kulp, Dale, and Starer, Amy (1995), "Bias in List-Assisted Telephone Samples," *Public Opinion Quarterly*, Volume 59: 218-235.

As indicated earlier, if there were more than one child in each age group, then one child in each age group was selected at random. If there were children in the household in only one age group, then two children were selected at random from that age group. In single-child households, the child was included in the sample.

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.

SLAITS Eligibility: Following the NIS eligibility question and, if applicable, the NIS portion of the interview, the respondents were asked whether the household income was above or below the amount determined to be 200% of poverty for a household of the composition indicated.

I. *Identification of Focal Child and Respondent*: Names and birthdates of household members under age 18 were rostered, and if the household included more than two children, the sampled children were selected. For this child (these children), gender and race/ethnicity data were 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 focal 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 focal 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 series was 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

Interviewing

An advance letter was mailed to households where a mailing address was identified using a Telematch process. The letter was translated into Spanish and printed with English on one side and Spanish on the other. Interviewing began on October 29, 1998, with interviewing conducted between 9:00 A.M. – 9:00 P.M. CST. Data collection, with 1,265 household interviews completed on March 19, 1999, yielded data on 686 children age 5 or younger and 1,323 children ages 6-17. In 56 of the households interviewed, two different adults were the "most knowledgeable adult" for the two children sampled. Interviews were completed with 34 second knowledgeable adults. In 2 cases, the interview was completed with a knowledgeable adult who was under age 18.

The number of calls made to complete an interview ranged from 1 to 98, with a mean of 11 calls and a median of 6 calls. There was some variation by household composition:

- For the cases with an NIS-eligible child, the mean number of calls was 9.3 compared with 11.8 for the other households.

- The mean number of calls to complete an English-language case was 11.9. The mean number of calls was 9.9 for cases completed in Spanish (despite the fact that the first call was often devoted to determining that the case required a bilingual interviewer).

- There is a negative relationship between the number of calls required to complete a case and the total number of household members, with two-member households requiring a mean of 14.6 calls and households with six or more members requiring an average of 10.6 calls.

- There is a positive relationship between household income and the number of calls required to complete the interview. For example, among those with an income of $12,500 or less (the quarter of respondents with the lowest reported incomes), the mean number of calls made was 8.6 calls (median of 4 calls); while the mean number of calls made to complete interviews with households whose income was $46,000 or greater (the quarter of respondents with the highest reported incomes) was 13.4 calls (median of 7 calls). Additionally, among households that did not report an income during the interview, the mean number of calls was even higher, at 13.9 calls (median of 8 calls).

- More calls were made to complete interviews in households with older children. In households where the youngest child was under age 6, an average of 9.9 calls (median of 5 calls) were made in order to complete an interview, compared with 13.2 calls (median of 8 calls) in a household where the youngest child was older then age 13.

Most of the cases pending (callbacks, breakoffs, refusals) at the end of the field period were those that had not yet been screened into or out of the NIS. That is, there was a comparatively smaller number of cases, 23%, that had been screened for NIS eligibility or, if the household contained an NIS-eligible child, had completed an NIS interview, but had not completed the SLAITS portion of the interview.

Interview Length

The interviews (including the SLAITS portion of the screener) averaged 30.3 minutes in length among NIS-ineligible households and 45.1 minutes (including the NIS portion) for NIS-eligible households. Among the NIS-ineligible households, this varied from a mean of 26.9 minutes for a household with one child compared with 32.5 minutes for a household with two or more children.

Other Languages

Five hundred sixty-one households required administration of the screener and/or interview in Spanish, but 210 of these households were screened out as not having children. Interviews were completed with 266 of the remaining households by a Spanish-speaking interviewer using the Spanish version of the questionnaire. In 63 households, a language other than English or Spanish was spoken; 24 of these households were screened out as not having children, and for 39 the SLAITS interview could not be attempted. (For five of these cases, the NIS portion of the interview was completed using the AT&T Language Line, but the SLAITS portion was not

included in this effort.)

Breakoffs and Refusals

Interviews were completed in 164 households that had initially refused to participate (13.1% of the completed interviews). Of the cases that were finalized as refusals, almost three-quarters (72.1%) refused or broke off prior to the SLAITS portion of the interview. Of the 162 refusal cases that had reached the SLAITS portion of the interview when the respondent ended the interview, the most common question on the interviewers' screens was the income eligibility question (22.9% of these SLAITS refusals). It is not clear whether the income question itself prompted a refusal since the respondent may have instead been reacting to the lengthy transition prior to the income question, and indeed this transition statement was on the interviewer's screen for an additional 17.6% of the SLAITS refusals.[2] The item producing the next most frequent refusals or breakoffs included the rostering of children's names (17.2%).

Reporting of Income

While some questions that may be considered sensitive by one respondent are innocuous to others, making identification of problematic questions difficult, questions about income were anticipated to be viewed as sensitive by a sizeable portion of the respondents. Income was included as part of the screener and then asked again, using the NIS income question series, near the end of the interview. The nonresponse rate to this question during the screener was 9.5% (6.8% don't know and 2.7% refusal). In order to gain some understanding of the respondents who do not provide a response during the screener, these respondents were interviewed despite not being able to screen them for income during the SLAITS screener.

There was some misclassification of income during the screening phase. Of the children whose households screened in as having an income at or below 200% of poverty, 13.8% actually were in a household with a family income above 200% of poverty, based on the respondents' answers to the more detailed income questions at the end of the interview. Similarly, 3.3% of the children whose households screened in with an income above 200% of poverty were actually in households with an income below 200% of poverty.

Response Rates

The interview completion rate was 87.5%, the screener completion rate was 87.2%, and the resolution rate was 90.7%.[3] Thus, the CASRO response rate was 69.9 to 70.2%, depending on the formula used for calculations. There was considerable variability across geographic areas in Texas, with the CASRO response rate varying from a low of 63.2% to 64.1% in Dallas County, to a high of 75.7%-77.9% in El Paso County.

---

2 This transition statement included the information regarding confidentiality, reiteration that participation is voluntary, direction that questions might be skipped, an estimated length of 20 minutes, and a reminder to the respondent that he or she might "end the interview at any time."

3 If the screener had been limited to determining the presence of children (without the final screening question regarding income), the screening completion rate would have been 91.5%.

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 household on that record. In households where two children were sampled, household-level information is repeated on the record for the second child.

Editing

Concurrent with the development of the CATI questionnaire for the 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 (such situations constitute hard edits). 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 (these were soft edits). 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. Each variable's range of permissible values 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. When necessary for later calculating sampling weights, some missing values were imputed (see next section). 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.

Imputation

When a request for an exact family income value was refused, or when the respondents did not know the answer, income was collected using a series of 16 unfolding questions (see questions W9Q02 - W9Q12a). This technique was used to reduce the proportion of "unknowns" that result from asking for an exact family income value. Still, the proportion of children whose parents or guardians either refused to answer the income questions or who didn't know their family income was 19.8 percent in Texas. Because the question content of the Texas CWBW Module is heavily weighted toward items correlated with socio-economic status (SES), it was deemed important to incorporate some measure of income into the sampling weights.

For all households where income values were missing, income was imputed using a hot-deck procedure within the households' geographic areas. That is, the data on income were borrowed from another record which closely matches the record for which income was missing. Imputed income (R_INCOME) is given on the public use data files along with a flag (I_FLAGIN) indicating which cases have an imputed income value.

Because race/ethnicity and age were also used in the sampling weights, these variables were imputed for children for whom information was missing. The variable I_FLAG23 indicates which cases have an imputed race/ethnicity for the child; I_FLAG45 indicates which cases have an imputed race/ethnicity for the mother; and I_FLAG6 indicates which cases have an imputed age for the child.

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, geographic information that would identify the specific IAP level in Texas has been dropped from the data set. The response categories for the race and ethnicity variables have been reduced to just four (non-Hispanic white, non-Hispanic Black, Hispanic, and other). Education level for adults has been recoded to four categories (up to 3 years high school, 4 years high school or high school graduate, 1-3 years college, 4 years college or more). Family income has been reported as 5 categories (<$15000, $15000 - $24999, $25000 - $34999, $35000 - $49999, $50000+). Individual income from employment was 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 two categories: public coverage (including Medicaid, Medicare, Medi-Gap, Indian Health Service, and military) and private coverage (including single-service plans and other kinds of insurance).

VI. Estimation

Sampling Weights

For producing population-based estimates of totals, percentages, ratios, and averages, a survey weight was attached to each respondent household and each respondent child. This weight combines: the base sampling weight, which reflects the varying probability of selection of a household and a child in the sample due to different rates of sampling; an adjustment for households that have multiple telephone numbers; an adjustment to compensate for unit nonresponse; a poststratification adjustment to a set of known population totals; and, finally, an adjustment to account for noncoverage of nontelephone households. The poststratification adjustment, in addition to adjusting for some undercoverage, also corrects any imbalance in the representation of important population subgroups in the sample, through the use of known control totals. This may also lead to a reduction in the variance of the estimates. In this section, we describe the procedure for determining both the household sampling weight and the child weight.

Base Sampling Weight

The first step in the weighting methodology was to assign a base sampling weight to each selected telephone number in each of the 5 IAP areas in Texas. The base sampling weight is similar to the one used in the NIS and was obtained by taking the ratio of the number of telephone numbers eligible for selection to the number of telephone numbers selected from the telephone banks and dialed. If the selected telephone number led to an eligible household within an IAP area, then the base sampling weight was attached to that household. An eligible household for this survey is a household with at least one child below the age of 18 years.

If the total number of telephone numbers in IAP area h is denoted by N_h and the number of telephone numbers dialed is n_h, then the base sampling weight for eligible households in IAP area h is given by

$$W_h ' \frac{N_h}{n_h}.$$

Adjustment for Multiple-Telephone Households

The base sampling weight of eligible households that have multiple voice-use telephone lines was adjusted to compensate for the higher probability of selection of these households. The adjustment divides the base sampling weight by the number of telephone lines in that household. Let the number of telephone lines in household i in IAP area h be A_{hi}. The adjusted weight is

$$W_{hi} = \frac{W_h}{A_{hi}}.$$

If the household has only one telephone line, then the adjusted weight is the same as the base sampling weight.

Unit Nonresponse Adjustment (1)

When a selected telephone number is called, three results are possible: (i) the number called is a household, (ii) the number called is not a working residential number (it could be a business or non-working number), or (iii) there is nonresponse to the screening attempt, and the status of the telephone number is unknown. The determination that the number called is not a working residential number, or that its status is unknown, will be made from the disposition codes assigned after several call attempts. Therefore, an adjustment was made to the sampling weight to account for nonresponse to the attempt to determine whether the telephone number selected is a household.

Let the number of known households out of the n_h telephone numbers called be n_{h1}. Let n_{h2} denote the number of non-residential numbers. Let n_{h3} denote the number of telephone numbers whose status is unknown. Let n_{h31} and n_{h32} denote the number of unknown household and non-household numbers out of the total of n_{h3}.

$$n_h = n_{h1} + n_{h2} + n_{h3}.$$

Let the estimated number of households out of the "status unknown" numbers be $\hat{n}_{h31}$. The estimate is obtained by applying to the number of unknown numbers the ratio of residential numbers to the total of residential and nonresidential numbers.

The first nonresponse adjustment is given by the ratio

$$\frac{n_{h1} + \hat{n}_{h31}}{n_{h1}}.$$

Unit Nonresponse Adjustment (2)

The households in the sample were screened for eligibility. A household was considered eligible to be included in the sample if it had children under the age of 18 years. Let the number of households known to be eligible out of n_{h1} be m_{h1}. Let m_{h2} denote the number of households that are ineligible. Let m_{h3} denote the number of households whose eligibility is unknown. Again, we

have to estimate the number of eligible households in this group.

$$n_{h1} = m_{h1} + m_{h2} + m_{h3}.$$

Let $\hat{m}_{h31}$ denote the estimated number of households out of m_{h3}. The second nonresponse adjustment is given by

$$\frac{m_{h1} + \hat{m}_{h31}}{m_{h1}}.$$

Unit Nonresponse Adjustment (3)

Out of the m_{h1} eligible households screened for poverty level, let the number of known eligible households with household income below 200% of the poverty level be q_{h1}. Let the number of known eligible households with household income equal to or above 200% of the poverty level be q_{h2}. Let q_{h3} denote the number of households that cannot be classified as being below or above 200% of poverty even after repeated attempts. We have

$$m_{h1} = q_{h1} + q_{h2} + q_{h3}.$$

Attempts were made to collect data from all eligible households with income below 200% of the poverty level. A subsample was selected from those households that are above 200% of poverty.

Let q'_{h2} denote the size of the subsample. Let q'_{h1} denote the number of households below 200% of poverty level for which complete data are available. Let q'_{h2} denote the number of households above 200% of poverty level for which complete data are available.

The nonresponse adjustment factors for the two income groups are (assuming that the proportions of unknown poverty group above and below 200% of poverty are in the same ratio as observed in the sample)

$$\frac{q_{h1} + \hat{q}_{h1}}{q'_{h1}} \qquad \text{and} \qquad \left(\frac{q_{h2} + \hat{q}_{h2}}{q'_{h2}}\right)$$

where $\hat{q}_{h1} = \frac{q_{h1}}{(q_{h1} + q_{h2})} q_{h3}$ and $\hat{q}_{h2} = \frac{q_{h2}}{(q_{h1} + q_{h2})} q_{h3}$.

Households classified as below 200% of poverty level at the screening stage may in fact turn out to have income above or equal to 200% of poverty level and vice versa. Therefore, for tabulations of children in households below 200% of poverty, contribution to the total estimates will come from two groups: those correctly classified as below 200% of poverty at the time of screening, and those incorrectly classified as above 200% of poverty but now correctly classified.

The nonresponse-adjusted base sampling weight for a respondent household with children and below 200% of poverty at the time of screening is given by

$$W^{(}_{hi} = W_{hi}\left(\frac{n_{h1}}{\hat{n}_{h31}}\right)\left(\frac{m_{h1}}{\hat{m}_{h31}}\right)\left(\frac{q_{h1}}{\hat{q}_{h1}}\right).$$

The weights for other households were computed similarly.

Child-Level Weight

Let the number of children in age group j in household i in IAP area h be C_{hij}. There are two age groups. Let the number of respondent children be c_{hij}. Then the weight for the respondent child in age group j is given by

$$W^{c}_{hij} = W^{(}_{hi}\frac{C_{hij}}{c_{hij}}.$$

Adjustment for Noncoverage of Nontelephone Households

Since SLAITS is a survey of telephone households, we have to adjust the sampling weights to account for households that did not have telephone service at the time of the survey and therefore had no chance of being selected for the survey. There is evidence to suggest that households with telephones at the time of the survey but which had experienced an interruption in service during the year are similar to households that did not have telephone service at the time of the survey. Therefore, the sampling weight adjustment was done using the data on interruption in telephone service obtained from the respondent households in the sample at the State level.

Adjustment to the weights for noncoverage of nontelephone households was accomplished through raking of child weights. Two telephone control totals were first determined. The first was the total number of children in Texas from households with telephones, but with no interruption in telephone service. The second was the total number of children from nontelephone households plus telephone households with an interruption in telephone service. These were used for raking (described in the next section). We first describe the procedure for establishing the telephone

control totals.

Step 1: We computed the weighted estimate of the number of children in the State in age group 0-17 coming from households with interruptions in telephone service from the survey data. Dividing this by the estimated total number of children in the State gave the estimated proportion of children coming from households with telephones but with interruption in telephone service in Texas. Let v denote this ratio.

Step 2: We estimated the proportion of children in the age group 0-17 coming from nontelephone households in Texas. This was based on the data from the CPS. Let P_{nt} denote this proportion.

Step 3: We obtained the total population of children in the age group 0-17 from the projections by single age years made by the Census and available on the Census website. Let this be denoted by N.

Step 4: We partitioned this total into subtotals. The first subtotal is the number of children in Texas coming from telephone households and the second subtotal is the number of children coming from nontelephone households. Let N_t denote the number of children from telephone households and N_{nt} denote the number of children from nontelephone households. We have

$$N_t = N\ (1-P_{nt}) \text{ and } N_{nt} = N\ P_{nt}.$$

Step 5: The number of children from telephone households was further split into two subtotals. The first is the number of children from telephone households, but without an interruption in telephone service. The second subtotal is the number of children from telephone households, but with an interruption in telephone service. Let N_{tnI} denote the first total of children without interruption and N_{tI} the number of children in Texas from telephone households with interruption. We have

$$N_{tnI} = N_t(1-?) \text{ and } N_{tI} = N_t v$$

Step 6: We formed two new control totals as follows.

N_{tnI} was the first control total and $N_o = N_{nt}+N_t v$ was the other control total.

Raking Procedure

As indicated earlier, child-level weights were raked imposing constraints in two dimensions or margins. The first margin consisted of the two telephone control totals derived earlier. The second margin consisted of poststratification cells created from race, gender, and age so that the sample frequency in each of the cells would be more than 30.

Final raked weights were consistent with the control totals in both margins. These weights are to be used for producing child-level estimates.

Variance Estimation and Hypothesis Testing

The data collected in SLAITS are obtained through a complex sample design involving both clustering and stratification. Because of the complex design, the direct application of standard statistical analysis methods for variance estimation and hypothesis testing may yield misleading results.

There are computer programs available which provide the capability of variance estimation for complex sample designs. The balanced repeated replication approach is utilized in &REPERR-&PSALMS-OSIRIS.IV to calculate the variance-covariance matrix. SESUDAAN, SURREGR, and SUDAAN are programs that calculated the variance-covariance matrix using the linearization approach (Taylor series expansion). In order to provide the user with the capability of estimating the complex sample variances in the SLAITS data using the above procedures, we have provided the Stratum Identifier and Primary Sampling Unit (PSU) codes on the data files. (The PSU for the Texas CWBW Module is the household; the data file includes records for one or two children per PSU.) These variables and the sample weights are necessary for the calculation of variances.

Even though the overall number of persons in this survey is sufficient for most statistical inference purposes, analyses of some rare responses and analyses of subclasses can lead to estimators that are unreliable. Consequently, these analyses require that the user pay particular attention to the coefficient of variation for the estimates of means, proportions, and totals. In addition, small sample sizes or a small number of PSU's used in the variance calculations may produce unstable estimates of the variances using the above computer programs.

Variance Estimation Using SUDAAN

This method requires no recoding of design variables and may be applicable to many complex survey sample design computer programs, but is statistically less efficient (and therefore more conservative) than some other methods because the PSU unit is treated as being sampled with replacement within the STRATUM unit. The data file needs to sorted only by STRATUM and PSU (i.e., household identifier) prior to invoking SUDAAN. The following SUDAAN design statements are used:

```
PROC  . . .  DESIGN = WR;
NEST STRATUM  PSU;
```

VII. 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, Texas, 1998-1999," 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 advice 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.