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

Health Module Pilot Test
Iowa and Washington, 1997

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. The first use of the SLAITS mechanism tested a health questionnaire in two States.

National Immunization Survey

The NIS is an ongoing Random-Digit-Dialing (RDD) survey designed to monitor on a quarterly basis vaccination coverage levels among children aged 19-35 months of age in 78 Immunization Action Plan (IAP) areas, consisting of the 50 states and 28 selected urban areas including the District of Columbia. The sum of the IAP areas represent the entire United States.

Integration of SLAITS with the NIS

Because the proportion of households with age-eligible children is small (about 4%), the NIS must screen 25 households to identify a single household with an age-eligible child. SLAITS was designed to take advantage of the cost-effective opportunity to use the large probability sample of telephone numbers to investigate emerging health and welfare issues.

Key Features of the 2 State Test

Important features of the first SLAITS test include the following:

- Use of NIS sampling frame during Q2/97 and Q3/97;
- Selection of Iowa and Washington States;
- No oversampling;
- A household-based questionnaire to collect data on each family member, including

health insurance coverage, access to care, utilization of health care services, health status (perceived and activity limitations), demographics, education, and income;
- 20-minute CATI interview;
- Respondent of age 18 or older (replaced by a more knowledgeable respondent if this person indicated that someone else would be better able to respond to the interview questions);
- Over 1,000 completed interviews in each of two states;
- One interview per family per household;
- An 8-week data collection period;
- NIS screener/interview completed prior to SLAITS portion of the interview;
- Interviews of NIS-eligible households as well as NIS-ineligible households;
- Non-English speaking households included through the use of a Spanish translation of the questionnaire administered by bilingual interviewers and the AT&T Language Line for other languages;
- Survey estimates include a weight adjustment strategy to allow for the representation of households without telephones.

The NIS collects data by computer assisted telephone interviews (CATI) from a random sample of households in the 78 Immunization Action Plan (IAP) areas. Although the target population for SLAITS in Iowa and Washington is all the households, the population from which the sample was drawn in each state was all households *with telephones*. Because the sampling frame for SLAITS in Iowa and Washington was the same as the one used for the NIS, a brief description of the NIS frame is given below.

NIS Sample

The NIS employs a list-assisted random digit dialing (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 (banks with no directory-listed residential number). These banks are deleted from the sampling frame. The Marketing Systems Group (MSG) list-assisted RDD system can screen out a portion of the nonworking numbers as a preliminary sample preparation activity. This component of MSG's product is called GENESYS-ID. In order to exclude business numbers, the sample of telephone numbers in the working banks was matched against a file of business numbers listed in the Yellow Pages. Then the remaining telephone numbers were dialed through an auto-dialer. The goal of the auto-dialer is to identify a nonworking number via nonworking-number intercept signal. Thus nonworking numbers can be eliminated from the sample at a very low cost. The process of eliminating zero banks and nonworking numbers yields a working residential 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, Waksberg, Kulp, & Starer, 1995 – Public Opinion Quarterly, v.59: 218-235).

SLAITS Sample

The telephone numbers for SLAITS in Iowa and Washington were selected by first identifying a certain number of replicates in each IAP area within each State. Iowa has only one IAP area whereas Washington has two IAP areas. The selection of telephone numbers was independent in each IAP area in Washington. The requirement was to have 1,000 completed household interviews in both Iowa and Washington. In Iowa, the number of replicates that was expected to yield 1,000 completed households interviews was calculated on the basis of the expected residential working telephone number rate and response rate. This number was then selected from the number of replicates available for NIS during the quarter in which SLAITS was conducted. In Washington, the sample of 1,000 households was first allocated to the two IAP areas in proportion to the total population in each IAP area given in the *NIS 1994 Annual Methodology Report*. The number of replicates required in each IAP area to yield the desired number of households was computed and then selected from the number available in each IAP area.

II. Questionnaire

The SLAITS questionnaire used in Washington and Iowa is a Health Module which incorporates questions from the National Health Interview Survey (NHIS), the Survey of Income and Program Participation (SIPP), and a short set of questions proposed by the Assistant Secretary for Planning and Evaluation (ASPE), Department of Health and Human Services. The seven sections of the questionnaire, following the screener, include:

I. *Household Composition*: This section includes a rostering of the household members, beginning with an owner or renter; and demographic information regarding each member (gender; relationship to the first-named person in the household; date of birth; presence of a child's biological, adoptive, step, or foster parent in the household; race/ethnicity; marital status; education, using the NIS education question; and educational degrees); and determination of whether the household includes multiple families.

II. *Health Care Access and Utilization*: This section includes information regarding the use of health care, such as the identification of whether each family member has a usual place where they go for health care and the type of place; incidence of hospital emergency room visits, overnight hospital stays, home health care, doctor office visits, and surgical procedures for each of the family members; inability to afford medical care, prescription medicines, mental health care or counseling, and dental care; and the length of time since the most recent dental visit.

III. *Health Status and Limitation of Activity*: This section includes information regarding the

impact of chronic physical, mental, or emotional health problems. The questions address limitations with respect to activities of daily living; working at a job or business; the kind or amount of work that can be performed; and for children under age 5, limitation in play activities. This section includes an evaluation of the general health status of each family member and whether there had been any change in the past 12 months.

IV. *Health Insurance*: This section addresses health insurance coverage for all members of the family. First, it was determined whether each family member was covered by health insurance or some other kind of health care plan and, if so, what type of health care coverage (private, Medicare, Medicaid, military, state-sponsored health plan, Indian Health Service, other government program, or some other insurance). For the family members without health insurance, information regarding who pays for medical care when needed, the last time they did have health insurance coverage, and the reasons they stopped being covered by health insurance was ascertained. For those covered by health insurance, it was determined whether there was any time during the past 12 months that they did not have any health insurance and for how long and why. This section also includes items regarding how much the family spent for medical care during the past 12 months.

V. *Sociodemographic Background*: This section asks about the family members' work status and, if not working, reasons for not working; whether health insurance was offered through the workplace and whether they were eligible to receive it. It also obtains information regarding how many days of work were missed because of illness or injury.

VI. *Income and Assets*: This section obtains information regarding sources of income including Social Security; Supplemental Security Income; Social Security Disability Income; Welfare, Temporary Assistance for Needy Families, or General Assistance; and Food Stamps. The NIS income question series is used to obtain total combined family income during the past 12 months.

VII. *Household Information*: This section includes items regarding birth outside of the United States, U.S. citizenship, additional telephone lines, and interruption in telephone service.

CATI Programming

Using the questionnaire programming system CASES, the questionnaire was converted into 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. There were two question series (on income and telephone lines) that were identical for the NIS and SLAITS portions of the interview. For respondents who had an NIS-eligible child, the NIS income series was asked near the end of the SLAITS portion of the interview. Questions regarding multiple telephone lines remained in the NIS portion for the NIS-eligible households, but were included at the end of the SLAITS interview for NIS-ineligible households.

Advance Letter

One of the procedures implemented to improve the quality of data was the use of advance introductory letters. Using a reverse directory to find addresses of all directory listed households enabled us to send introductory letters to 40% of the households in the sample. Selected households were informed of the importance of the survey and assured that all identifiable information obtained in the interview would be held in strict confidence. These advance letters were mailed beginning on May 22. During the telephone interview, if respondents indicated that they had not received or did not remember the letter, the critical portions were read to them.

Interviewing

Interviewing began on June 5 and ended on August 14 resulting in 1,021 household interviews in Iowa and 1,068 in Washington. Data collection includes completion of interviews with additional families in multiple-family households on hard copy. The interview was completed with a household member age 18 or older. If the household respondent was unable to adequately answer the questions, another adult household member was identified and interviewed either during that telephone call or a later one.

The number of calls made to complete an interview ranged from 1 to 54, with a mean of 6.5 calls and a median of 3 calls. There was some variation by state and household composition. In Iowa, the mean number of calls was 4.9 compared with 7.1 in Washington. For the cases with a NIS-eligible child, the mean number of calls was 7.7 compared with 5.7 for the other households. The mean number of calls to households with someone under age 65 was 6.2 compared with 3.8 for households where all members were age 65 or older.

Three-quarters (74%) of the cases pending (callbacks, breakoffs, refusals) at the end of the field period were still at the NIS stage of the interview. That is, there was a comparatively smaller number of cases (167 in Iowa and 158 in Washington) 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.

Breakoffs and Refusals

Of the cases that were finalized as refusals, the most common location that a refusal occurred was prior to determining NIS eligibility status (258 cases). The next most common location was after the introduction to the SLAITS portion of the interview (192 cases), followed by refusals after the household members had been enumerated (37 cases). Interviews were completed in 170 households that had initially refused to participate.

Response Rates

The interview completion rates among households were 75.9% in Iowa and 75.1% in Washington State.

Taking into account the resolution rate and enumeration rate as well as the interview completion rate, the final response rates were 67.95% in Iowa and 66.22% in Washington State.

Interview Length

The interviews averaged 20.2 minutes in length among NIS-ineligible households and 34 minutes (including the NIS portion) for NIS-eligible households.

Multiple-Family Households

Of the households where a roster was completed (2,389), 70 (2.9%) households (31 in Iowa, or 2.6%; and 39 in Washington State, or 3.2%) were identified as having multiple families. In 57 of these households, there was 1 additional family; in 8 of these households, 2 additional families, and in 5 of these households, 3 additional families. Among these 88 additional families, 75 had one member, 9 had 2 members, 3 had 3 members, and 1 additional family had 5 members. Interviews were completed with 49 additional families in these multiple-family households using a hard copy version of the questionnaire; this resulted in completed interviews for all the families in 36 of these 70 multiple-family households.

Other Languages

The questionnaire programmed in Spanish following translation by one translator from English to Spanish and then the Spanish version was translated back into English by a different translator. Spanish was determined to be the language spoken in 35 households (12 in Iowa, 23 in Washington). Interviews were completed by a Spanish-speaking interviewer using the Spanish version of the questionnaire with 31 of these households.

In 28 households (7 in Iowa, 21 in Washington), a language other than English or Spanish was spoken. For these cases, an interview was attempted using the AT&T Language Line. This service includes the translation of the questions, as read by the interviewer, by an AT&T Language Line operator to the respondent and then translation of the responses. Interviews were completed with 12 of these households.

III. Data Files

A SAS file containing the household-, family-, and member-level data variables was created for each of the two states.

Editing

Concurrent with the development of the CATI questionnaire for the SLAITS 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 persons 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 40-year old respondent reports coming to live in the USA in 1959, but the interviewer attempts to enter 1949, 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. One variable (NEW.250) was removed from the data set due to an error in skip patterns. Other variables that were missing responses for unknown reasons were left blank. 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

Income for SLAITS in Iowa and Washington was collected using a series of nine unfolding questions (see questions FIN12 - FIN18). This technique was used to reduce the proportion of "unknowns" that result from asking for an exact family income value. The proportion of respondents either refusing to answer the income questions or who didn't know their family income was 11.6 percent in Washington and 12.9 percent in Iowa. Because the question content of SLAITS 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.

Imputed income (IMPINC10) is given on the public use data files along with a flag (INC_FLAG) indicating which cases have an imputed income value. Income was imputed for each state independently using a randomized split sample approach where one sample was designated the developmental sample and the second sample the confirmation sample. A family-level OLS regression model was developed with the final model including the following variables: telephone exchange median family income; someone in family receive retirement income; someone in family receive assistance; more than one telephone; marital status (married, not married); education; family size; gender; have private health insurance; race (black + white, other); and self-assessed health status. The confirmation sample indicated that these models accounted for 69% of the variance in Iowa residents' reported income and for 66% of the variance in Washington residents' reported income. It can be further noted that the model correctly predicted the income category for approximately 32% of the confirmation sample, and the predicted income category was close (plus or minus one category) for approximately 73% of the sample.

Because gender was also used in the sampling weights, gender was imputed for the six individuals (1 in Iowa and 5 in Washington) for whom information on gender was missing. The information was imputed using a hot-deck procedure within an IAP area, income group, and age group. The variable GEN_FLAG indicates which cases have an imputed gender value.

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 Washington has been dropped from the data set. For both states, the response categories for the race variable have been reduced to just two (white and other), and H260 (specific Hispanic ancestry) has been dropped. Age has been reported in categorical form, using six categories (<5, 5-17, 18-24, 25-44, 45-64, and 65+). Education level has been recoded to four categories (less than high school, high school graduate, some college, and college graduate). Income has also been reported as 4 categories (<$15000, $15000 - $29999, $30000 - $49999, $50000+).

In addition, risk of inadvertant 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, data about participation in welfare, general assistance, food stamps, and other public assistance programs have been combined into one variable: PUBASST. This variable is reported at the family level, such that positive responses indicate that at least one member of the family has received the benefit.

Releasing data about Medicare and Medicaid eligibility also provides a risk of disclosure for those non-traditional enrollees (adults in Medicaid, non-elderly 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, a new variable (other public health insurance: OTHPUBHI) was created to represent these non-traditional enrollees, as well as persons reporting coverage through the Indian Health Service and other government programs.

IV. Estimation

For producing population-based estimates of totals and percentages, a survey weight was attached to each individual respondent in SLAITS. This weight combines the base sampling weight which reflects the probability of selection of an individual in the sample, 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 use of the sampling weights is REQUIRED to provide unbiased State estimates as well as to accurately assess the sampling error of statistics based on the survey data. When the appropriate weights are used, the data are representative of the total State population.

Base Sampling Weight

The sampling unit in SLAITS was a telephone number. The first step in the weighting methodology was to assign a base sampling weight to each selected telephone number in an IAP area. 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 divided by the number of telephone numbers selected from the telephone banks. If the selected telephone number was for a household within the State (either Washington or Iowa), then the base sampling weight was attached to that household. If the total number of telephone numbers in the hth IAP area is denoted by N_h and the number of telephone numbers selected is n_h , the base sampling weight for households in the hth IAP area is given by

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

Multiple-Telephone Households

The second step was to adjust the base sampling weight of households that have multiple voice-use telephone numbers. This adjustment was needed to compensate for the higher probability of selection of households with two or more telephone lines. The adjustment was done by simply dividing the base sampling weight of the household by the number of telephone lines in that household. Let the number of telephone lines in the ith household in the hth IAP area be A_{hi}. The weight attached to that household is given by

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

If the household had only one telephone line, then the two weights were the same.

Unit Nonresponse Adjustment

For unit nonresponse adjustment, the telephone numbers that were called were classified into three categories: (1) The number called was a household, (2) the number called was not a residential working number (it was either a business or non-working number) and (3) the status of the telephone number was unknown. The first adjustment was to account for the fact that some telephone numbers with an unknown status could be households. This was done based on the nonresponse disposition codes assigned to the telephone numbers selected and called in the survey (and resulting in a specific unresolved category). The adjustment is explained below.

Let the number of telephone numbers called in the hth IAP area be n_h. Let the number of known households out of n_h called be n_{h1}. Let the number of business and non-working numbers be n_{h2}. Let the number of telephone numbers that have unknown status be n_{h3}. We have

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

The n_{h3} telephone numbers that have unknown status in Iowa were classified into four groups depending on whether a telephone number was in an exchange located in an MSA or a non-MSA and whether it was directory-listed or not. The telephone company business offices were called and a sample number of telephone numbers were classified as residential or non-residential according to the business office information. Adjustment factors were developed for each type of disposition code in both directory-listed and non-directory-listed categories. A similar procedure was followed for Washington State IAP areas except that in one of the IAP areas unresolved numbers were grouped into 8 categories, instead of 4, depending on whether an unresolved telephone number was directory-listed or not, whether it was located in an MSA or non-MSA, and whether the percent of college graduates was more or less than 20%. Again adjustment factors were developed for each type of disposition code for both directory-listed and non-directory-listed telephone numbers. The adjustment factors, when applied to the total number of unresolved numbers in each category, give the estimated number of households. Let this number be $\hat{n}_{h31}$. The nonresponse adjustment factor is given by the ratio

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

The second unit nonresponse adjustment was to account for nonrespondent households. Let n_{h11} be the number of households for which we have complete data and n_{h12} are nonrespondents. We have

$$n_{h1} = n_{h11} + n_{h12}.$$

The nonresponse adjustment to the data collection attempt is given by

$$\frac{n_{h1}}{n_{h11}}.$$

The adjusted sampling weight for the ith households in the hth IAP area is given by

$$W^{(}_{hi} = W_{hi}\frac{n_{h1}\%\hat{n}_{h31}}{n_{h1}}\ \frac{n_{h1}}{n_{h11}}.$$

Simple Poststratification

Several poststratification adjustments to the nonresponse-adjusted base sampling weight were tried before choosing an appropriate poststratification adjustment and the adjustment for noncoverage of non-telephone households. The procedure for determining the final weight which was used to produce the weighted counts and percentages is described below. The poststratification adjustment is described as being carried out in several steps although in actual practice all the steps were done at the same time.

In Iowa, the sample of individual respondents was poststratified by nonHispanic White and Other race categories. The nonHispanic White group was further stratified by sex and eight age categories within male and female groups, and the Other race category was stratified by sex, thus creating 18 poststrata. Population totals were obtained for each of the 18 categories using the 1997 Census estimates and the sampling weights of individuals within each of these 18 poststrata were adjusted such that the sum of the weights agreed with the control totals. A similar procedure to derive the poststratification weights in Washington State was adopted except that the sample was divided into four race/ethnicity categories, nonHispanic White, Hispanic, Asian/Pacific Islander and Others, creating 22 poststrata. Household income was considered an important variable and therefore, income with ten categories was introduced as an additional poststratification variable. Income was imputed for all individuals who were in the missing income category. The control totals for the ten categories of income both in Iowa and Washington State were obtained by averaging the estimates in these categories from the 1995 and 1996 Current Population Survey. Also, since the age variable was not used to poststratify the Other group in Iowa and Washington, the age variable with eight categories was introduced as an additional poststratification variable. An 18 x 18 matrix with the control totals along the margins was created for Iowa. Similarly, a 22 x18 matrix was created for Washington with the 22 control totals obtained earlier along the rows and the 18 income and age control totals along the columns of the matrix. The weights in each cell were raked such that the sum of the weights along the columns and rows agreed with the control totals.

The final step in adjusting the weight was to account for noncoverage of nontelephone households. This was accomplished by retaining all the poststratification categories described previously and adding some new categories and splitting some existing categories. The control totals for each of the eight age categories were split into two totals; one being the total number of individuals having public insurance and the other not having public insurance. Similar splits were obtained for the ten income categories. Two race and two sex categories were introduced as additional poststratification variables in Iowa with their control totals split into two parts, one having public insurance and the other not having public insurance. In Washington State, four race categories and two sex categories were introduced, again splitting the control totals into two; one having public insurance and the other not having public

insurance. If the number of completed interviews in the public insurance part of a category was very small, we did not split the category. Weights in each cell of the matrices were raked such that the sum of the weights along the columns and rows agreed with the control totals in the margins. The resulting weights in both Iowa and Washington were used to produce the estimates of totals and percentages.

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. 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 prior to invoking SUDAAN. The following SUDAAN design statements are used:

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

V. 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, [State Name(s)], 1997," 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.