

National Immunization Survey

A User's Guide for the 1995 Public-Use Data File

Centers for Disease Control and Prevention

**National Immunization Program
and
National Center for Health Statistics**

**Prepared by Abt Associates Inc.
May 2002**

Table of Contents

1.	Introduction.....	4
2.	Sample Design	6
	The NIS RDD Sample	7
	The NIS Provider Record Check Study	8
	Summary of Data Collection.....	9
	Informed Consent, Security, and Confidentiality of Information.....	12
3.	Content of NIS Questionnaires	13
	Content of the NIS Household Questionnaire	13
	Content of the Immunization History Questionnaire	16
4.	Data Preparation and Processing Procedures.....	16
	Data Preparation.....	17
	Limitations of Data Editing Procedures.....	20
	Variable-Naming Conventions	22
	Missing-Value Codes	23
	Imputation for Item Nonresponse	24
	Vaccine-Specific Recoding of Verbatim Responses	24
	Composite Variables.....	25
	Subsets of the Data.....	26
	Confidentiality and Disclosure Avoidance	27
5.	Quality Control and Quality Assurance Procedures	27
6.	Sampling Weights.....	29
	Adjusted Base Sampling Weight	30
	Adjustment for Interview Nonresponse	30
	Adjustment for Households That Do Not Have Telephones	31
	Adjustment for Provider Nonresponse.....	33
7.	Analytic and Reporting Guidelines.....	36
	Key Variables.....	36
	Use of the NIS Sampling Weights	40
	Estimation and Analysis	41
8.	Summary Tables	43
9.	Citations for NIS Data.....	44
10.	References	44

Appendices

Appendix A: Glossary of Abbreviations and Terms

Appendix B: NIS Household Questionnaire

Appendix C: NIS Provider Questionnaire

Appendix D: Summary Statistics for Sampling Weights by IAP Area

Appendix E: Disposition of Child with respect to Provider Record Check for NIS, Q1/1995 to Q4/1995

Appendix F: Examples of the Use of SUDAAN to Estimate Vaccination Coverage Rates and Their Standard Errors

Appendix G: Table of Contents and Alphabetical Index of Variables from National Immunization Survey 1995 Public-Use Data File: Documentation, Code Book and Frequencies

Appendix H: Summary Tables

1. Introduction

In 1992 the Childhood Immunization Initiative (CII) (CDC 1994) was established to 1) improve the delivery of vaccines to children; 2) reduce the cost of vaccines for parents; 3) enhance awareness, partnerships, and community participation; 4) improve vaccinations and their use; and 5) monitor vaccination coverage and occurrences of disease. Subsequently the Healthy People 2000 and 2010 objectives established the goal of having at least 90% of 2-year-old children fully vaccinated with the recommended schedule of vaccines. To fulfill the CII mandate of monitoring vaccination coverage and marking progress toward achieving those goals, the National Immunization Survey (NIS) has been implemented by the National Immunization Program and the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC), and its contractor, Abt Associates Inc.

The target population for the NIS is children aged 19 to 35 months living in the United States at the time of the interview. The official coverage estimates reported from the NIS are rates of being up-to-date with respect to the recommended numbers of doses of all recommended vaccines (CDC 1995). These vaccines and their recommended numbers of doses are: diphtheria and tetanus toxoids and pertussis vaccine (DTP), 4 doses; poliovirus vaccine (polio), 3 doses; measles-containing vaccine (MCV), 1 dose; *Haemophilus influenzae* type b vaccine (Hib), 3 doses; and hepatitis B vaccine (Hep B), 3 doses. In addition to these vaccines, interest focuses on coverage rates for vaccine series, including the 4:3:1:3 series (4 DTP, 3 polio, 1 MCV, and 3 Hib). The NIS collects data on each of these vaccines.

The NIS uses a random-digit-dialing (RDD) telephone survey to identify households containing children in the target age range and interview an adult who is most knowledgeable about the child's vaccinations. With the consent of the child's parent or guardian, the NIS also contacts (by mail) the child's health care providers to request information on vaccinations from the child's medical records.

Samples of telephone numbers are drawn independently, for each calendar quarter, within 78 Immunization Action Plan (IAP) areas. Of the 78 IAP areas, 28 (including the District of Columbia) are urban areas. The remaining 50 are either an entire state or a "rest of state" IAP area (where the state contains one or more urban IAP areas). This design makes it possible to produce annualized estimates of vaccination coverage levels within each of the 78 IAP areas with a specified degree of precision (a coefficient of variation of no more than 5%). Further, by using the same data collection methodology and survey instruments in all IAP areas, the NIS produces vaccination coverage levels that are comparable among IAP areas and over time.

For the 1995 NIS the RDD interviews of households began in January 1995 and ended in February 1996, and provider data collection extended from around March 1995 to around April 1996. A total sample of approximately 2 million telephone numbers yielded household interviews for 31,997 children, and 16,183 of those children had provider data that were adequate to determine whether the child was up-to-date with respect to the recommended immunization schedule. The 1995 NIS public-use file (PUF) contains data for the 31,997 children with completed household interviews (and more extensive data for children with

provider data). Published tables of estimates of vaccination coverage for 1995 are available on the National Immunization Program (NIP) Web site, <http://www.cdc.gov/nip/coverage>, and are discussed in an *MMWR* article (CDC 1997).

The accompanying code book (*National Immunization Survey 1995 Public-Use Data File: Documentation, Code Book and Frequencies*) documents the contents of the 1995 NIS public-use data file. For reference Appendix G reproduces the table of contents and the alphabetical index of variables from the code book.

Additional information on the NIS is available at:

www.cdc.gov/nip/coverage
www.cdc.gov/nis/
www.nisabt.org

For additional information on the NIS data file, please contact the NCHS staff:

Data Dissemination Branch, NCHS
6525 Belcrest Road, Room 1000
Hyattsville, MD 20782
Phone: 301-458-INFO (301-458-4636)
E-mail: nchsquery@cdc.gov
Internet: <http://www.cdc.gov/nchs/>

2. Sample Design

The NIS uses two phases of data collection to obtain vaccination information for a large national probability sample of young children: a random-digit-dialing survey designed to identify households with children 19 to 35 months of age, followed by the Provider Record Check study (PRC), which obtains provider-reported vaccination histories for these children.

This section gives a summary of these two phases of data collection. Other descriptions of the sample design are given by Ezzati-Rice et al. (1995), Zell et al. (2000), and Smith et al. (2001a).

The NIS RDD Sample

The NIS RDD sampling phase uses independent quarterly samples of telephone numbers in the 78 IAP areas. Table H.1 (in Appendix H) lists the 78 IAP areas by state and shows the estimated number of children living in each state and IAP area in 1995.

The NIS uses the list-assisted method of random-digit dialing (Lepkowski 1988). This method selects a random sample of telephone numbers from “banks” of 100 consecutive telephone numbers (e.g., 617-495-0000 to 617-495-0099) that contain one or more directory-listed residential telephone numbers. The sampling frame of telephone numbers is updated each quarter in order to include new telephone exchanges and area codes. Although the number of cellular telephone users in the U.S. has increased rapidly, most households continue to maintain land-line telephone service. Also, most cellular telephone users pay for incoming calls. Therefore, the NIS sampling frame excluded cellular telephone exchanges in 1995.

The target sample size for each IAP area was 110 children with completed telephone interviews per quarter (440 for the year). Approximately 51% of children with completed telephone interviews had adequate provider data. The phrase “adequate provider data”

means that sufficient vaccination history information was obtained from the providers to determine whether the child is up-to-date with respect to the recommended vaccination schedule. The percentage of children with adequate provider data varies among the IAP areas.

The design and implementation of the NIS sample involve four procedures. First, statistical models predict the number of sample telephone numbers needed in each IAP area to meet a target number of interviews (Buckley et al. 1998). Second, the sample for an IAP area is divided into random subsamples called replicates. By administering the sample release on a replicate-by-replicate basis, it is possible to spread the interviews for each IAP area evenly across the entire calendar quarter. Third, an automated procedure eliminates a portion of the nonworking and nonresidential telephone numbers from the sample before the interviewers dial them (Battaglia et al. 1995b). Fourth, the sample telephone numbers are matched with a national database of directory-listed residential telephone numbers in order to obtain usable mailing addresses for as many sample households as possible. To promote participation in the NIS, an advance letter is sent to these addresses approximately two weeks prior to the RDD interview.

The NIS Provider Record Check Study

At the end of the NIS RDD interview, consent to contact the child's vaccination providers is requested from the parent/guardian. When verbal consent is obtained, those providers are

mailed an immunization history questionnaire (IHQ). This mail survey portion of the NIS is the Provider Record Check Study (PRCS).

The IHQ is sent by mail to vaccination providers with instructions to mail or fax the questionnaire back upon completion. Two weeks later, a thank you/reminder postcard is sent to each provider. If no response has been received, another questionnaire packet is mailed five weeks after the initial mailing. Finally, seven weeks after the initial mailing, a telephone call is made to providers who have still not responded, to remind and encourage them to complete the form and either mail or fax the information back. In some instances, provider-reported vaccination histories are accepted over the phone. The data from the IHQs are entered, cleaned, edited, and merged with the household information from the RDD survey to produce a child-level record.

Summary of Data Collection

Table 1 presents selected operational results of NIS data collection for calendar year 1995. Children who were 19 to 35 months of age during 1995 were born from February 1992 to May 1994. The original sample (in replicates that were released for use) consisted of 1,917,474 telephone numbers. Of those, 407,259 numbers were eliminated by the automated procedure as nonworking or nonresidential numbers. The remaining 1,623,857 telephone numbers were called to identify 929,066 households, as shown in Rows 3 and 6 of Table 1. Among the identified households, 853,536 (96.4%) were successfully screened for age-eligible children. Of these, 819,825 did not contain an age-eligible child, and 33,711 (3.9%)

contained one or more age-eligible children. Among these households 31,520 (93.5%) completed the NIS household RDD interview.

A standard approach for measuring response rates for RDD surveys, known as the CASRO household response rate, has been defined by the Council of American Survey Research Organizations (Frankel 1983). In 1995 the CASRO household response rate (Row 11) was 87.1%. The CASRO response rate equals the product of the resolution rate (96.5%, Row 5) the screening completion rate (96.4%, Row 7) and the interview completion rate among eligible households (93.5%, Row 10). The resolution rate is the percentage of the total phone numbers called that are classifiable as nonworking, nonresidential, or residential. The screening completion rate is the percentage of known households that are successfully screened for the presence of age-eligible children. The interview completion rate is the percentage of households with one or more age-eligible children that complete the NIS RDD interview.

Row 12 of Table 1 shows that 31,997 age-eligible children had completed RDD interviews. Rows 13 through 16 of Table 1 give results for the PRC phase. Specifically, Row 13 gives the rate of obtaining verbal consent from household respondents to contact their children's vaccination providers – approximately 84% in 1995. The number of IHQs that were mailed to vaccination providers exceeds the number of completed child interviews, because some children have more than one vaccination provider. In 1995 the mean number of vaccination providers identified for a child was 1.32.

Among vaccination providers who were mailed an IHQ, approximately 61% returned the questionnaire or other information pertaining to the child's vaccination history. Among the children with completed household RDD interviews 16,183 (50.6%) had adequate vaccination histories returned by their vaccination provider(s). The other 49.4% of children lacked adequate provider data for a variety of reasons, such as the parent did not give consent to contact providers, or the providers did not have medical records for the child.

For each IAP area and each state Table H.1 shows the number of children with completed RDD interviews and the number of children with adequate provider data.

Informed Consent, Security, and Confidentiality of Information

The Screener Introduction, the Advance Letter, and the Oral Consent assure the respondent of the confidentiality of his/her responses and the voluntary nature of the survey. Informed consent is obtained from the respondent (generally the parent or guardian of the child) to participate in the household interview and also (at the end of the interview) to contact the child's vaccination providers.

Information in the NIS is collected and processed under high security. To ensure privacy of the respondents and confidentiality of sensitive information, NCHS has established standards for release of data from all NCHS surveys. All CDC staff and contractor staff involved with the NIS sign the NCHS confidentiality agreement and follow instructions to prevent disclosure.

All information in the NIS is collected under strict confidentiality and can be used only for research purposes [Section 308(d) of the Public Health Service Act, 42 U.S. Code 242m(d), and the Privacy Act of 1974 (5 U.S. Code 552a)]. Prior to the public release, the contents of the PUF go through an extensive review by the NCHS Disclosure Review Board to protect confidentiality of the participants as well as the data.

3. Content of NIS Questionnaires

This section describes the questionnaires used in the 1995 NIS telephone interview of households and in the NIS PRC survey. The confidentiality of respondents and their data is required by Section 308(d) of the Public Health Service Act [42 U.S. Code 242m(d)].

Content of the NIS Household Questionnaire

The Computer-Assisted Telephone Interview (CATI) questionnaire used in the RDD phase of NIS data collection (Appendix B) consists of two parts: a screener to identify households with children aged 19 to 35 months and an interview portion. The questionnaire is modeled on the Immunization Supplement to the National Health Interview Survey (NHIS) (NCHS 1999). The NIS CATI questionnaire has been translated into Spanish, and the AT&T Language Line is used for real-time translation into many other languages (Wall et al. 1995). Table 2 summarizes the content of each section of the 1995 NIS household interview.

Table 2: Content of the 1995 NIS Household Interview

Screener	Screening questions to determine eligibility, roster of eligible children, availability of shot records
Section MR	Most-knowledgeable-respondent callback questions
Section SR	Shot-record callback questions
Section A	Vaccination history, asked if shot records are available
Section B	Vaccination history, asked if shot records are not available
Section C	Demographic and socioeconomic questions
Section D	Provider information and request for consent to contact the eligible child's vaccination providers
Section NR	Nonresponse questions

In the screener the purpose of the survey is explained to the respondent, and the household is screened to determine whether it contains any children 19 to 35 months of age. If the household has an eligible child, the respondent is asked whether he/she is the most knowledgeable person (MKP) for the child's vaccination history. If the respondent indicates that another person in the household is more knowledgeable, the interviewer asks to speak to him or her at that time. If that person is unavailable to be interviewed, the interview proceeds to Section MR, the name of the MKP is recorded, and a "callback" is scheduled for a later date.

Also during the screener the person being interviewed is asked whether he/she has a written record (shot card) of the child's vaccination history, and whether it is easily accessible. If the shot card is available, the respondent is asked to provide information directly from it in

Section A. If the child does not have a shot card, the interview proceeds with Section B, which asks the respondent to recall from memory information about the child's vaccinations. If the child has a shot card but it is not easily accessible, the interview proceeds to Section SR. In this section the interviewer makes an appointment to call at a later date, when the shot card will be available, and also gathers general information about the child's immunization history.

Section C obtains information that includes the relationship of the respondent to the child, the race of the child, the race of the mother, household income and educational attainment of the mother of the child, and other information on the socioeconomic characteristics of the household and its eligible children. This section is asked of all respondents upon completion of Section A, B, or SR.

At the conclusion of the NIS household interview, consent is requested to contact the child's vaccination providers (Section D). If verbal consent is obtained, identifying information (name, address, and telephone number) on the vaccination provider(s) is requested, as well as the full names of the child and the respondent, so that NIS personnel can contact the providers and identify the child whose immunization information the NIS is requesting. When verbal consent and sufficient identifying information are obtained, the IHQ is mailed to the child's vaccination provider(s). *No changes were made to the NIS CATI questionnaire in 1995.*

Content of the Immunization History Questionnaire

The IHQ (Appendix C) is designed to be simple and brief, to minimize burden on the providers and to encourage participation in the survey. It consists of two pages. Page 1 includes space for a label that gives the child's name and birth date and the full name of the parent or guardian. Page 1 also includes a grid for recording dates of vaccinations. The columns of the grid correspond to recommended vaccines, and an additional column is available for recording other vaccines. Page 2 of the IHQ contains several questions about the facility and vaccination provider (for example, whether the facility is public or private). *No changes were made to the IHQ in 1995.*

4. Data Preparation and Processing Procedures

The household data collection and provider data collection in the NIS incorporate extensive data preparation and processing procedures. During the household interview the CATI system makes many edits as the interviewer enters the data. After the completion of interviewing for a quarter, post-CATI editing and data cleaning produce a final interview data file. The editing of the provider data begins with a manual review of returned IHQs, data entry of the IHQs, and cleaning of the provider data file. If a provider reported a different date of birth, gender, or name for the child, a check of the IHQ is made to determine whether the IHQ was filled out for the correct child. After the provider data are merged with the household interview data, and responses from multiple providers for a child are consolidated into a child-level data record, the editing continues. The end product is an

analytic file containing household and provider data for use in estimating vaccination coverage.

Data Preparation

The editing and cleaning of NIS data involve several steps. First, the CATI system incorporates an automatic editing process. Further cleaning and editing take place in a post-CATI clean-up stage, involving an extensive review of data values, crosschecks, and the recoding of verbatim responses for race, ethnicity, and vaccinations. The next step involves the creation of numerous composite variables. Finally, provider data are cleaned in a separate step. After these steps have been completed, imputations are performed for item nonresponse on selected variables, and weights are calculated. The procedures and rules of the National Health Interview Survey served as the standard in all stages of data editing and cleaning.

Editing in the CATI System

The CATI software checks consistency across data elements and does not allow interviewers to enter invalid values. Catching potential errors early increases the efficiency of post-survey data cleaning and processing.

The CATI system makes a number of edits as an interviewer enters data. These edits correct data entry errors that can be reconciled while the respondent is on the telephone; they focus,

in particular, on items critical to the conduct of the study, such as those that determine a child's eligibility (e.g., date of birth). To the extent possible without making the CATI system overly complicated, out-of-range and inconsistent responses produce a warning screen, allowing the interviewer to correct errors as they occur.

A CATI system cannot simultaneously incorporate every possible type of error check and maximize system performance. To reconcile this trade-off, post-CATI edits are used to resolve problems that do not require access to the respondent, as well as unanticipated logic problems that appear in the data.

Post-CATI Edits

The post-CATI editing process produces final, cleaned data files for each quarter. The steps in this process, implemented after all data collection activities for a quarter are completed, are described below.

Initial Post-CATI Edits and File Creation

After the completion of interviewing each quarter, the raw data are extracted from the CATI data system and used to create two files: the Sample File and the Interview File. The Sample File contains one record for each sample telephone number. It contains summary information for telephone numbers and households. The Interview File contains one record for each eligible sample child. It contains all vaccination data that the household reported for the child.

Following the creation of these files, a preliminary analysis of each file identifies out-of-range values and extraneous codes. The first check verifies the eligibility status of children, based on date of birth and date of interview. Once the required corrections are verified, the invalid values are replaced with either an appropriate data value or a missing-value code.

Frequency Review

After the pre-programmed edits are run, frequency distributions of all variables in each file are produced and reviewed. Each variable's range of values is examined for any invalid values or unusual distributions. If blank values exist for a variable, they are checked to see whether they are allowable and whether they occur in excessive numbers. Any problems are investigated and corrected as appropriate.

File Crosschecks

Crosscheck programs make sure that cases exist across files in a consistent manner. Specifically, checks ensure that each case in the Interview File is also present in the Sample File and that each case in the Sample File was released to the CATI center. Checks also ensure that no duplicate households exist in the Sample File and no duplicate children exist in the Interview File.

When all of these checks have been performed, the final quarterly Interview File is created. Programmers and statisticians then create composite variables for each child. Sampling weights (described in Section 6) are added to each record.

Editing of Provider Data

Six to eight weeks after the close of household data collection for a quarter, the collection of Immunization History Questionnaires from providers typically ends. The data from the hard-copy questionnaires are entered and independently re-entered to provide 100% verification. The Provider Data File is cleaned, in a similar fashion to the household data, for out-of-range values and consistency. A computer program back-codes all “other shot” verbatim responses into the proper vaccine category (e.g., Energix B counts as Hep B, and Tetramune counts as DTP and Hib). These translations come from a file that contains all such verbatim responses ever encountered in the NIS. Also, the Provider File is checked for duplicate records, and exact duplicates are removed from the file. If the IHQ contains a date of birth of the child, gender of the child or child name that differs from the household interview, the IHQ is examined to see whether it may have been filled out for the wrong child. IHQs that appear to have been filled out for the wrong child are removed from the provider database. When a child has data from more than one IHQ, decision rules are applied to produce the most complete picture of the child’s immunization history.

Once these data have been cleaned, they are combined with the household interview data. Information from up to five providers can be added to a child’s record.

Limitations of Data Editing Procedures

Although data editing procedures were used for the 1995 NIS, the data user should be aware that some inconsistent data may remain in the public-use file. The variables that indicate

whether a child is up-to-date on each vaccine or series (on which the estimates of vaccination coverage are based) are derived from provider-reported data. Hence the household-reported vaccination dates (from interviews conducted with a shot card) are not edited for discrepancies beyond the built-in checks in the CATI system.

The NIS does not recontact households or providers to attempt to reconcile potential discrepancies in provider-reported vaccination dates or to resolve date-of-birth reporting errors. However, beginning with the 1999 NIS, the provider-reported data were manually reviewed and edited to correct specific reporting errors. The *National Immunization Survey: Guide to Quality Control Procedures* discusses the editing procedures in more detail. The provider-data edits that were added in 1999 are not reflected in the 1995 NIS data. A small number of children will therefore have provider-reported vaccination dates that contain potential discrepancies. Among children in the 1995 NIS PUF with adequate provider data, 1.2% have one or more provider-reported vaccination dates listed before the date of birth of the child, 4.2% have vaccination dates less than or equal to 14 days apart, and 1.8% have a vaccination other than hepatitis B reported as administered from 0-37 days of life. The section on Subsets of the Data (below) includes additional information related to the first dose of hepatitis B vaccine, which for the 1995 NIS was often given early in life. Overall, even with these minor limitations, the NIS is a rich source of data for assessment of up-to-date status and age-appropriate immunization.

Variable-Naming Conventions

To facilitate access to the contents of the PUF, the names of variables adhere to the SAS (Version 6.12) convention of having no more than 8 characters, and they follow a systematic pattern as much as possible. The code book for the PUF groups the variables into nine broad categories according to the source of the data (household or providers) and the content of the variable (see Appendix G).

The household report of vaccinations received by the child is used to create household up-to-date indicator variables. The names of these variables begin with FULL. For example, FULL_HEP indicates whether the child has received three or more hepatitis B vaccinations. Additional household up-to-date variables combine each vaccine with use of a shot card. The names of these variables begin with C_. For example, C_HEP has five values, corresponding to up-to-date on hepatitis B from a shot card, not up-to-date on hepatitis B from a shot card, up-to-date on hepatitis B not from a shot card, not up-to-date on hepatitis B not from a shot card, and vaccination status on hepatitis B indeterminate.

The provider data from the IHQs are used to create numerous child-level composite variables, as described below. The names of the variables giving the number of doses received for each vaccine begin with P_NUM. For example, P_NUMHEP gives the number of doses of hepatitis B vaccine according to the provider data. An up-to-date indicator variable also exists for each vaccine, and these variables begin with P_UTD. For example, P_UTDHEP indicates whether the child received 3 or more doses of hepatitis B vaccine.

The provider data are also used to form variables for age in days and age in months at time of vaccination. For age in days and age in months, either 4 or 8 variables are created, depending on the vaccine. The variables for age in months end with n_AGE, where n is the dose number. For example, HEP1_AGE to HEP8_AGE give age in months for 8 possible doses of hepatitis B vaccine. Similarly, for age in days at vaccination, the variables start with D and end with the dose number. For example, DHEPB1 to DHEPB8 give age in days for 8 possible doses of hepatitis B vaccine.

Missing-Value Codes

The missing-value codes for household variables are 6 and 96 for DON'T KNOW and 7 and 97 for REFUSED. Some household variables may also contain blanks, if the question was not asked. The variables developed from the IHQ generally do not have specific missing-value codes. For example, if a provider failed to answer the question on types of care provided, the response category variables for that question would be blank. For provider-reported vaccination dates before the date of birth of the child, the age in months and age in days variables are recoded to **-1**. For provider-reported vaccination dates after 36 months of age, the age in months variables are recoded to **40**. For the corresponding provider-reported vaccination dates after 1125 days of age, the age in days variables are recoded to **1210**. The official published estimates of vaccination coverage include these vaccination dates in the count of vaccines received by a child.

Imputation for Item Nonresponse

The NIS uses imputation primarily to replace missing values on selected socioeconomic and demographic variables collected in the household survey. A sequential hot-deck method is used to assign imputed values (Cox 1980). Each imputation cell has at least four donors.

The Notes section of the code book identifies variables that contain imputed values. These variables include maternal education, Hispanic origin, race, and race/ethnicity.

Vaccine-Specific Recoding of Verbatim Responses

During the household interview, respondents are given the option to report vaccinations in addition to, or instead of, the categories specifically read to them. These verbatim responses are entered into the CATI system by the interviewer and stored in the Interview File. They are reviewed in the post-CATI editing process in order to reclassify the responses into the listed categories, where possible. NIP personnel manually review the verbatim responses and determine to which category or categories (for combination shots), if any, each should be recoded. Once the recoding has been completed, a quality control review ensures that the responses were correctly recoded and are consistent with one another.

Composite Variables

A number of composite variables (constructed from basic variables) are created and included in the NIS PUF. Composite variables assist users and data analysts by eliminating duplication of effort and making NIS data easier to use.

Since the initial years of NIS data collection, the household composite variables have included up-to-date status on individual vaccinations, race of child and mother, household income, and up-to-date status on several vaccination series. Many of these composite household variables are included in the NIS PUF. Table 3 lists some of the key demographic variables and their categories.

Table 3: Key Demographic Composite Variables

AGEGRP – age category of child	19-23 months 24-29 months 30-35 months
RACEKIDR – race/ethnicity of child	Hispanic White, nonHispanic Black, nonHispanic All other, nonHispanic
SEX – gender of child	Male Female
EDUC1 – education of the mother	<12 years 12 years >12 years, not a college graduate College graduate
MARITAL – marital status of mother	Widowed, divorced or separated Never married Currently married Deceased Unknown
INCPOV1R – poverty status	At or above poverty level Below poverty level Not determined

The composite race variables in the 1995 PUF contain three categories: white, black and all other races. The “all other races” category includes American Indian, Asian, Alaska Native, Native Hawaiian, Pacific Islander, and other races. If more than one race was selected during the administration of the race questions, the respondent was asked to select the race that best characterizes the child/mother. The 1995 PUF uses these questions to assign each child and mother to a single race category.

The provider data from the IHQs are used to create numerous child-level composite variables. The most important variables give the number of doses received for each type of vaccine. Up-to-date indicator variables are created for each individual vaccine and for several vaccine series. Another set of variables gives age in days at time of vaccination. For each dose of a vaccine, the age in days is constructed from the date of birth of the child and the date of the shot. Corresponding variables give exact age in months at time of vaccination.

The IHQs also contain information on provider characteristics. This information is used to create composite variables related to provider facility type (PROV_FAC), and types of care offered by the provider (CARTYP1 to CARTYP5).

Subsets of the Data

The NIS PUF contains data for all children aged 19 to 35 months who have a completed household (RDD) interview. An interview is considered complete if the respondent

answered either Section A or Section B of the questionnaire. As explained in Section 6, each child with a completed household interview is assigned a weight (HY_WGT) for use in estimation.

The NIS uses the provider-reported vaccination histories to form the estimates of vaccination coverage, because the provider data are considered much more accurate. Thus, the most important subset of the data consists of children with adequate provider data. For these children one or more providers returned the IHQ, and the vaccination information reported by those providers is sufficient to determine whether the child is up-to-date on the recommended vaccinations. **As discussed in Section 7, the PDAT variable identifies the children with adequate provider data (PDAT=1). These children have a separate weight (W0) that should be used to form estimates of vaccination coverage.**

Confidentiality and Disclosure Avoidance

To prevent identification of participants in the NIS and the resulting disclosure of information, certain items from the questionnaires are not included in the PUF. In addition, some of the released variables are top- or bottom-coded, or their categories are collapsed.

5. Quality Control and Quality Assurance Procedures

A major contributor to the quality of the NIS data is its sample management system, which manages 312 RDD samples annually (78 IAP areas times 4 quarters) and uses 20

performance measures to track their progress toward completion. Important aspects of the quality assurance program for the RDD component of the NIS include on-line interviewer monitoring; on-line look-ups in topic-oriented databases integrated with the CATI system, including names, addresses and telephone numbers of vaccination providers; and automated range-edits and consistency checks. These and other quality assurance procedures contribute to a reduction in the total cost of the data collection, by minimizing interviewer labor and overall burden to respondents. Khare et al. (2000), Khare et al. (2001), and the *National Immunization Survey: Guide to Quality Control Procedures* discuss the procedures in more detail.

The quality assurance procedures of the PRC component follow a proven methodology documented by Dillman (1978). The most critical quality assurance activities occur during post-processing of the returned questionnaires or vaccination records. All returned IHQs are examined to identify and correct any obvious errors prior to data entry and then key-entered with 100% verification. The National Immunization Program additionally has conducted a manual quality assurance review of 10% of forms returned by providers. Resulting error rates for the edit process are estimated to be less than 1%.

Some special conditions apply to the first dose of hepatitis B, which is typically given at 0 to 7 days. The count of vaccinations for a specific vaccine is based on the number of unique vaccination dates reported by the child's provider(s). For a very small percentage of children the provider may indicate on the IHQ that the child received hepatitis B at birth but they do not record a vaccination date. Because no date is given, this dose is not included in the count

of hepatitis B vaccinations for these children, resulting in a slight underestimation of hepatitis B vaccination coverage.

6. Sampling Weights

Each of the two stages of data collection results in a sampling weight for the children who have data at that stage. The RDD sampling weights (HY_WGT) permit analyses of data from children with completed household interviews. Each child with adequate provider data (the subset on which official estimates of vaccination coverage are based) has a “partial-nonresponse-adjusted sampling weight” (W0).

A sampling weight may be interpreted as the approximate number of children in the target population that the child in the sample represents. Thus, for example, the sum of the sampling weights of children who are up-to-date (on a particular vaccine or series of vaccines) yields an estimate of the total number of children in the target population who are up-to-date. Dividing this sum by the total of the sampling weights for all children gives an estimate of the corresponding vaccination coverage rate.

This section describes how these weights are developed and adjusted so as to achieve an accurate representation of the target population. The weights reflect each child’s probability of being selected into the sample; and the adjustments take into account the number of telephone lines in the household, nonresponse to the household interview, noncoverage of households that do not have telephones, and nonresponse by providers.

Adjusted Base Sampling Weight

In each quarterly NIS sample, each child with a completed RDD interview receives a base sampling weight. This weight is equal to the total number of telephone numbers in the sampling frame for the IAP area divided by the total number of telephone numbers that were randomly sampled from that sampling frame during that quarter. Because households with multiple telephone lines have a greater chance of being sampled, each child's base sampling weight is adjusted by dividing it by the total number of residential telephone lines reported in the household (up to a maximum of 3).

Adjustment for Interview Nonresponse

Nonresponse occurs in population-based surveys when respondents refuse to participate or are not available at the time of the interview. Thus, the sum of the adjusted base sampling weights of children with completed RDD interviews will underestimate the size of the target population in the IAP area, because some sampled households containing age-eligible children do not complete the RDD interview. As a result, the adjusted base sampling weights must be further adjusted so that they more accurately reflect the number of children in the target population that each sampled child with a completed RDD interview represents.

Some sampled households with age-eligible children fail to complete the RDD interview because of unit nonresponse: some telephone numbers are never determined to be residential despite multiple call attempts, some households cannot be determined to have age-eligible

children, and some households with age-eligible children do not complete the RDD interview. To compensate for these three types of unit nonresponse, the sampling weights of children with a completed RDD interview are adjusted to account for the estimated number of age-eligible children in households whose telephone numbers are never determined to be residential, the estimated number of age-eligible children in households that fail to complete the screening interview, and the number of identified age-eligible children for whom the RDD interview is not completed. Each of these adjustments is carried out within IAP areas by forming weighting cells based on the residential directory-listed status of the sample telephone number and socioeconomic and demographic characteristics of the IAP area's telephone exchanges (e.g., 4 weighting cells formed from directory-listed versus non-directory-listed telephone number by telephone exchanges with 75% or higher white population versus telephone exchanges with less than 75% white population).

Because the quarterly interview-nonresponse-adjusted base sampling weights pertain to the entire target population and because annualized vaccination coverage estimates are obtained from data for four consecutive quarters, the adjusted base sampling weights are divided by 4 when the data from the four quarters are combined.

Adjustment for Households That Do Not Have Telephones

The NIS sampling frame includes only households that have telephones. Because the target population consists of all children 19 to 35 months of age living in households regardless of whether they have telephones, the interview-nonresponse-adjusted base sampling weights

need to be adjusted to compensate for the noncoverage of children living in households without telephones. Although national telephone coverage for age-eligible children is estimated to be 90%, telephone coverage is known to be as low as 76% in some IAP areas. Further, data from the NHIS, which samples both “telephone” and “nontelephone” households, indicate that children living in households without telephones have significantly lower vaccination coverage. Thus, the adjustment to the sampling weights to compensate for noncoverage of nontelephone households may be particularly important in IAP areas in which the percentage of households that have telephones is relatively low.

In order to reduce the impact of this potential bias, two separate adjustments to sampling weights are made. In the first adjustment, the weighted distributions of “poststratification” variables, which are known to be strongly associated with variation in vaccination coverage rates, are adjusted to agree with those obtained from Vital Statistics (NCHS 1993) compiled by the National Center for Health Statistics (NCHS). The poststratification variables are race/ethnicity of the child’s mother, the level of educational attainment of the child’s mother, and the age of the child. Because the Vital Statistics data give the counts of all live births in the U.S., regardless of whether the household has telephone service, this adjustment corrects in part for underrepresentation of children who belong to households that are less likely to have telephones (typified by racial/ethnic minorities or mothers with low educational attainment).

The second adjustment for nontelephone households in 1995 depends on whether a sample child is up-to-date on the 4:3:1 vaccination series and also on two other factors: the IAP-area-

specific proportion of children that live in households that do not have telephones, as estimated from the 1990 Census and the Current Population Survey (Bureau of Labor Statistics 2000) for each combination of levels of the poststratification variables described above, and the ratio of the national 4:3:1 vaccination coverage rate among children living in nontelephone households to the national 4:3:1 vaccination coverage rate among children living in telephone households, as estimated using data for major race/ethnicity groups from the NHIS.

For children belonging to a specific race/ethnicity group, the adjustment to the sampling weight is larger for children who are not 4:3:1 up-to-date than for children who are 4:3:1 up-to-date when: the percentage of children living in nontelephone households in the IAP area is large and the estimated national 4:3:1 vaccination coverage rate among children living in nontelephone households is less than the estimated national 4:3:1 vaccination coverage rate among children living in telephone households. In this situation the adjustment for households that do not have telephones tends to reduce estimated vaccination coverage rates slightly. A further description is given by Battaglia et al. (1995a).

The base sampling weights after adjustment for multiple residential telephones, unit nonresponse, and nontelephone coverage constitute the “RDD sampling weights.”

Adjustment for Provider Nonresponse

Among the 31,997 children with a completed RDD interview, 16,183 (50.6%) had adequate provider data. The 15,814 (49.4%) children for whom an RDD interview was completed but

adequate provider data were not obtained are “partial nonresponders” because they provide a partial response to the NIS as a whole.

Empirical results suggest that children with adequate provider data have characteristics that are believed to be associated with a greater likelihood of being up-to-date, compared to partial nonresponders. Specifically, children with adequate provider response are more likely to live in households that have higher total family income, to have a white mother, and to live outside a central city of a Metropolitan Statistical Area. Also, a partial nonresponder is less likely to live in the state where the mother resided when the child was born and less likely to have a parent/guardian who could locate a shot card. Both of these factors indicate a potential lack of continuity of health care, and are associated with lower vaccination rates (Coronado et al. 2000). If no adjustment is made to the RDD sampling weights to account for these differences, estimated vaccination coverage rates may be biased.

To reduce potential bias in estimated vaccination coverage estimates attributable to partial nonresponse, a “weighting-class adjustment” is used in each IAP area (Brick and Kalton 1996). This adjustment involves two steps. In the first step, sampled children are classified according to the quintile of their estimated probabilities of having adequate provider data. In the statistical literature these probabilities are called response propensities (Rosenbaum and Rubin 1983, 1984; Rosenbaum 1987). Children who have similar response propensities will also be similar with respect to variables that are strongly associated with the probability of having adequate provider data. In this important respect, children in each class are comparable. Because of this comparability, any subsample of children in a class may

represent all of the children in the class. Therefore, the weighting-class adjustment uses the children with adequate provider data to represent all of the children in the class.

In the second step of the weighting-class adjustment, within each class, an adjustment factor redistributes the RDD sample weights of the partial nonresponders among the children who have adequate provider data. These revised RDD sampling weights of children with adequate provider data are “partial-nonresponse-adjusted RDD sampling weights” (W0). Because of the comparability of children within each weighting class, any estimate that uses data only from the children with adequate provider data, along with their partial-nonresponse-adjusted RDD sampling weights, will have less bias attributable to differences between children with adequate provider data and partial nonresponders. Smith et al. (2001b) describe the development of this approach in more detail. Appendix D summarizes the distribution of the sampling weights (HY_WGT and W0) in each IAP area.

In 1999 the CDC adopted this nonresponse-adjustment methodology for producing all public-use files and for calculating estimates of vaccination coverage for 1998 (and subsequent years). It replaced the approach that had been used to calculate estimates for 1995. Thus, estimates calculated from the 1995 PUF may differ slightly from those in the *MMWR* article (CDC 1997). Estimates from the 1995 PUF do agree with those published at the NIP website (<http://www.cdc.gov/nip/coverage>), in the column labeled Tables, and also shown in Table H.7 in Appendix H of this User’s Guide, which are based on the new methodology. More information can be found at the NIP website (<http://www.cdc.gov/nip/coverage/NIS/analysis-methodologies.htm>).

7. Analytic and Reporting Guidelines

The NIS PUF can be used to produce national, state and IAP area estimates of vaccination coverage rates. Information in the data file can be used to calculate standard errors of the vaccination coverage rates that reflect the complex sample design of the NIS. The file includes IAP area and state identifiers (ITRUEIAP and STATE). The sample is stratified by the 78 IAP areas, and the IAP area identifier and the coded household identifier (SEQNUMHH) are key variables for obtaining standard errors for IAP area, state and national estimates of vaccination coverage rates. Demographic and socioeconomic variables in the file can be used to obtain national vaccination coverage rates for subgroups of the population. Data users should, however, be aware that estimates for such subgroups at the state or IAP area level will generally have large standard errors because of the small sample sizes. The NCHS standard for precision of subgroup estimates is that the ratio of the standard error to the estimate should be less than or equal to 30%, and each analytic cell should contain at least 30 respondents.

Key Variables

The variables in the NIS PUF fall into two major categories: 1) variables that apply to all children with completed household interviews and 2) variables that apply only to children with adequate provider data (i.e., PDAT=1). Variables in the first group include the household report of vaccinations received by the child, and various demographic and socioeconomic characteristics of the child, the mother and the household. Because of

reporting and recall errors, the household report of vaccinations is not used to produce vaccination coverage rates. As discussed below, the provider report of vaccinations received by the child is used to produce vaccination coverage rates.

Table 4 lists variables that are commonly used in analyses or for published estimates of vaccination coverage.

The SEQNUMC variable is the unique child identifier. Key geographic variables include IAP area (ITRUEIAP), state (STATE), and Census Region (REGION). Key demographic variables include race/ethnicity category of the child (RACEKIDR), age category of the child (AGEGRP), and marital status category of the mother (MARITAL). Key socioeconomic variables include education category of mother (EDUC1), and poverty status (INCPOV1R).

Selecting children with PDAT equal to 1 identifies children with adequate provider data (DISPCODE = 1 to 6 or 8 to 11). Children who do not have provider data (DISPCODE = MISSING) or who have provider data that are not adequate to determine the up-to-date vaccination status of the child (DISPCODE = 7) have PDAT equal to 2. (Appendix E gives the definition of the values of DISPCODE.) The NIS PUF contains many variables constructed from the provider data. One set of variables indicates the number of doses the child received for each of the vaccines. For example, P_NUMDTP indicates the number of doses of DTP. It counts all DTP-containing vaccines, including DTP, DTaP, DT and DTP-Hib.

Table 4: NIS Variables That Are Commonly Used in Analyses or for Published Estimates

ID variables	
SEQNUMC – unique child ID variable	
Geographic variables	
ITRUEIAP – IAP area	
STATE – state FIPS code	
REGION – Census Region	Northeast Midwest South West
Child demographic variables	
AGEGRP – age category of child	19-23 months 24-29 months 30-35 months
RACEKIDR – race/ethnicity of child	Hispanic White, nonHispanic Black, nonHispanic All other, nonHispanic
SEX – gender of child	Male Female
Mother demographic variables	
EDUC1 – education of the mother	<12 years 12 years >12 years, not a college graduate College graduate
MARITAL – marital status of mother	Widowed, divorced or separated Never married Currently married Deceased Unknown
RACEMOMR – race/ethnicity of mother	Hispanic White, nonHispanic Black, nonHispanic All other, nonHispanic
Poverty variables	
INCPOV1R – poverty status	At or above poverty level Below poverty level Not determined

Presence of provider data variables	
PDAT – adequate provider data indicator	Yes No
Number of provider-reported doses of vaccine variables	
P_NUMDTP – total number of DT/DTP/DTaP doses	
P_NUMPOL – total number of Polio doses	
P_NUMMMR – total number of MCV doses	
P_NUMHIB – total number of Hib doses	
P_NUMHEP – total number of Hep B doses	
Provider characteristic variables	
PROV_FAC – provider facility type	All public facilities All hospital facilities All private facilities All military/other facilities Mixed types Unknown
CARTYP1 to CARTYP5 – types of services offered by child’s provider(s)	All providers Some but not all providers No providers/unknown

Both the individual vaccines and the vaccine series have up-to-date indicator variables. For example, PUTD4313 is an indicator variable for whether the child has 4 or more DTP vaccinations, 3 or more polio vaccinations, 1 or more measles-containing vaccinations (MCV), and 3 or more Hib vaccinations. Section 4 discusses the naming conventions for these variables.

The NIS PUF includes a set of variables for age in days at each vaccination. These variables can be used to examine age at vaccination, vaccination spacing intervals, and age-appropriate immunization. Another set of variables gives age in months at time of vaccination. These variables can be used to determine, for example, whether a child received at least four DTP

vaccinations by the age of 19 months. Section 4 discusses the naming conventions for these variables.

The final key set of provider variables relates to characteristics of the provider: provider facility type (PROV_FAC), and type of care offered by the provider (CARTYP1 to CARTYP5).

Use of the NIS Sampling Weights

The NIS PUF contains two child-level weights. The HY_WGT variable gives the household weight for each child. It should be used to form estimates from the children with completed household interviews. This weight reflects the stratified sample design and also adjusts for unit nonresponse, for poststratification to population control totals, and for the exclusion of nontelephone children from the NIS. **The weight variable that applies to children with adequate provider data is W0. This weight should be used to form estimates of vaccination coverage. Each child with adequate provider data (PDAT = 1) has a value of W0.**

The NIS PUF does not contain any provider-level weights. The NIS does not sample providers directly; rather, they are included in the survey through the children they vaccinate. A user of the NIS PUF should not attempt provider-level analyses (e.g., estimate the percentage of providers in the U.S. that are private providers), because the NIS sample was not designed for that purpose.

Estimation and Analysis

Estimating Vaccination Coverage Rates

Vaccination coverage rates are ratio estimates, as described by the statistical literature on methods for complex sample surveys. Because of the adjustment to the sampling weights for partial nonresponse, statistical analyses require only data from children with adequate provider data (PDAT = 1), along with their partial-nonresponse-adjusted sampling weights (W0). To summarize the statistical methodology by which vaccination coverage rates and their standard errors are obtained from these data, let Y_{hij} be an indicator, for the j th child with adequate provider data in the i th sampled household in the h th stratum (IAP area) of the NIS sampling design, that is equal to 1 if the child is up-to-date according the provider data and 0 otherwise. Also, let W_{hij} denote the value of W0 for this child. Then, letting

$\hat{Y}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} Y_{hij}$ and $\hat{T}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij}$, the national estimator of the vaccination coverage rate

may be expressed as

$$\hat{q} = \frac{\sum_{h=1}^L \hat{Y}_h}{\sum_{h=1}^L \hat{T}_h}$$

where L denotes the number of strata (the 78 IAP areas), n_h denotes the number of sampled households containing children with adequate provider data in the h th IAP area, and m_{hi} denotes the number of age-eligible children with adequate provider data in the i th household in the h th IAP area.

