**CALIFORNIA HEALTH INTERVIEW SURVEY**

**CHIS 2003 METHODOLOGY SERIES**

# REPORT 3

# DATA PROCESSING PROCEDURES

*December 2005*

*This report was prepared for the California Health Interview Survey by John Rauch and W. Sherman Edwards of Westat, Inc*


*www.chis.ucla.edu*

This report describes the data processing and editing procedures for CHIS 2003 that were performed by Westat. It discusses standard data editing procedures and addresses the steps taken for ensuring data quality. It also presents discussions on special procedures of editing and coding of the geography, industry and occupation, and race and ethnicity survey items.

**Suggested citation:**

California Health Interview Survey. *CHIS 2003 Methodology Series: Report 3 – Data Processing Procedures*. Los Angeles, CA: UCLA Center for Health Policy Research, 2005.



The California Health Interview Survey is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. Funding for CHIS 2003 came from multiple sources: the California Department of Health Services, the California Endowment, the National Cancer Institute, the Robert Wood Johnson Foundation, the California Office of the Patient Advocate, the U. S. Centers for Disease Control and Prevention, Kaiser Permanente, the Alameda County Health Care Agency, and L.A. Care Health Plan.

# PREFACE

*Data Processing Procedures* is the third report in a series of methodological reports describing the 2003 California Health Interview Survey (CHIS 2003). The other reports are listed below. This report describes the data processing procedures that took place at Westat. It does not include the additional processing procedures performed later by UCLA. Please check the CHIS website (www.chis.ucla.edu) for availability of reports on the data processing procedures at UCLA.

CHIS is a collaborative project of the University of California, Los Angeles (UCLA) Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. Westat was responsible for data collection and the preparation of five methodological reports from the 2003 survey. The survey examines public health and health care access issues in California. The telephone survey is the largest state health survey ever undertaken in the United States. The plan is to monitor these issues and examine changes over time by conducting surveys in the future.

## Methodological Reports

The first five methodological reports for CHIS 2003 are as follows:

- Report 1: Sample Design;
- Report 2: Data Collection Methods;
- Report 3: Data Processing Procedures;
- Report 4: Response Rates; and
- Report 5: Weighting and Variance Estimation.

The reports are interrelated and contain many references to each other. For ease of presentation, the references are simply labeled by the report numbers given above.

This report describes the data processing and editing procedures for CHIS 2003. One chapter details the data editing procedures and addresses the steps taken for ensuring data quality. Delivery of the final data sets is also discussed. Another chapter presents information about the geographic and the

industry and occupation coding. The next chapter describes how the race and ethnicity survey items were coded for CHIS.

## TABLE OF CONTENTS


| Chapter | | | Page |
|---|---|---|---|
| | PREFACE | | i |
| 1 | CHIS 2003 DESIGN AND METHODOLOGY SUMMARY | | 1-1 |
| | 1.1 | Overview | 1-1 |
| | 1.2 | Sample Design Objectives | 1-2 |
| | 1.3 | Data Collection | 1-3 |
| | 1.4 | Response Rate | 1-4 |
| | 1.5 | Weighting the Sample | 1-8 |
| | 1.6 | Imputation Methods | 1-9 |
| | 1.7 | Methodology Report Series | 1-10 |
| 2 | DATA EDITING PROCEDURES | | 2-1 |
| | 2.1 | Resolving Problem Cases | 2-2 |
| | 2.2 | Interviewer Comments | 2-3 |
| | 2.3 | Coding with Text Strings | 2-4 |
| | 2.4 | Verifying Data Updates | 2-7 |
| | 2.5 | Data Conversion and Delivery | 2-8 |
| 3 | GEOGRAPHIC AND INDUSTRY AND OCCUPATION CODING | | 3-1 |
| | 3.1 | Geographic Coding | 3-1 |
| | 3.2 | Industry and Occupation Coding | 3-11 |
| 4 | RACE AND ETHNICITY CODING | | 4-1 |
| | 4.1 | Coding Procedures | 4-2 |


### List of Tables


| Table | | |
|---|---|---|
| 1-1 | California county and county group strata used in the CHIS 2003 sample design | 1-2 |
| 1-2 | Number of completed CHIS 2003 interviews by type of sample, instrument | 1-4 |
| 1-3 | CHIS 2003 Survey topic areas by instrument | 1-5 |

## TABLE OF CONTENTS

List of Tables (continued)


| Table | | Page |
|---|---|---|
| 2-1 | Response codes added to CHIS 2003 ..... | 2-6 |
| 3-1 | Zip Codes within sampling stratum ..... | 3-3 |
| 3-2 | Final distribution of adult extended completed cases by self-reported and original sampling stratum ..... | 3-10 |

# 1. CHIS 2003 DESIGN AND METHODOLOGY SUMMARY

## 1.1 Overview

The California Health Interview Survey (CHIS) is a population-based random-digit dial telephone survey of California's population that is conducted every two years. First conducted in 2001, CHIS is the largest health survey ever conducted in any state and one of the largest health surveys in the nation. CHIS is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. CHIS collects extensive information for all age groups on health status, health conditions, health-related behaviors, health insurance coverage, access to health care services, and other health and development issues.

The CHIS sample is designed to provide population-based estimates for most California counties, all major ethnic groups, and several ethnic subgroups. The sample is designed to meet and optimize two goals: provide estimates for large- and medium-sized population counties in the state, and for groups of the smallest population counties; and provide statewide estimates for California's overall population, its major race/ethnic groups, as well as for several Asian ethnic groups. The resulting CHIS sample is representative of California's non-institutionalized population living in households.

This series of reports describes the methods used in collecting data for the 2003 California Health Interview Survey (CHIS 2003). CHIS 2001 is described in a series of methodology reports.[1] These reports describe the second CHIS data collection cycle, which was conducted between August 2003 and February 2004.

CHIS data and results are used extensively by many State agencies, local public health agencies and organizations, federal agencies, advocacy and community organizations and agencies, foundations, and researchers. They use these data in their own analyses and publications to assess public health and health care needs, to develop health policies, and to develop and advocate policies to meet those needs.

[1] California Health Interview Survey, CHIS 2001 Methodology Series: Report 1 - Sample Design, Report 2 – Data Collection Methods, Report 3 – Data Processing Procedures, Report 4 – Response Rates, and Report 5 – Weighting and Variance Estimation, Los Angeles, CA: UCLA Center for Health Policy Research, 2002.

## 1.2 Sample Design Objectives

The CHIS sample is designed to meet two objectives: (1) provide estimates for counties and groupings of counties with populations of 100,000 or more; and (2) provide estimates for California's overall population and its larger race/ethnic groups, as well as for several smaller ethnic groups. To achieve these objectives, CHIS relied on a multi-stage sample design. First, the state was divided into 41 geographic sampling strata, including 33 single-county strata and 8 groups that included the 25 other counties. Second, within each geographic stratum, households were selected through random-digit dial (RDD), and within each household, an adult (age 18 and over) respondent was randomly selected. In addition, in those households with adolescents (ages 12-17) and/or children (under age 12), one adolescent was randomly selected for interview and one child was randomly selected and the most knowledgeable parent of the child interviewed.

Table 1-1 shows the 41 sampling strata (i.e., counties and groups of counties that were identified in the sample design as domains for which separate estimates would be produced). A sufficient amount of sample was allocated to each of these domains to support the first sample design objective. These strata were also used for the CHIS 2001 sample; because of funding limitations, the sample sizes allocated to most strata for CHIS 2003 were smaller than in 2001.

Table 1-1. California county and county group strata used in the CHIS 2003 sample design

| | | |
|---|---|---|
| 1. Los Angeles | 15. San Joaquin | 29. El Dorado |
| 2. San Diego | 16. Sonoma | 30. Imperial |
| 3. Orange | 17. Stanislaus | 31. Napa |
| 4. Santa Clara | 18. Santa Barbara | 32. Kings |
| 5. San Bernardino | 19. Solano | 33. Madera |
| 6. Riverside | 20. Tulare | 34. Monterey, San Benito |
| 7. Alameda | 21. Santa Cruz | 35. Del Norte, Humboldt |
| 8. Sacramento | 22. Marin | 36. Lassen, Modoc, Siskiyou, Trinity |
| 9. Contra Costa | 23. San Luis Obispo | 37. Lake, Mendocino |
| 10. Fresno | 24. Placer | 38. Colusa, Glen, Tehama |
| 11. San Francisco | 25. Merced | 39. Sutter, Yuba |
| 12. Ventura | 26. Butte | 40. Plumas, Nevada, Sierra |
| 13. San Mateo | 27. Shasta | 41. Alpine, Amador, Calaveras, Inyo, |
| 14. Kern | 28. Yolo | Mariposa, Mono, Tuolumne |

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

The samples in Los Angeles and Alameda Counties were enhanced with additional funding to allow sub-county geographic estimates, in Los Angeles at the Service Planning Area (SPA) level and in Alameda for the cities of Oakland and Hayward as well as the remainder of the county. These samples were implemented with and incorporated into the original statewide RDD sample.

To accomplish the second objective, larger sample sizes were allocated to the more urban counties where a significant portion of the state's Latino, African American and Asian ethnic populations reside. To increase the precision of the estimates for Koreans and Vietnamese, areas with relatively high concentrations of these groups were sampled at higher rates; these geographic samples were supplemented by phone numbers for group-specific surnames drawn from listed telephone directories to increase the sample size and precision of the estimates for these two groups.

## 1.3 Data Collection

To capture the rich diversity of the California population, interviews were conducted in five languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, and Korean. These languages were chosen based on research that identified the languages that would cover the largest number of Californians in the CHIS sample that either did not speak English or did not speak English well enough to otherwise participate.

Westat, a private firm that specializes in statistical research and large-scale sample surveys, conducted the CHIS 2003 data collection. Westat staff interviewed one randomly selected adult in each sampled household. In those households with children (under age 12) or adolescents (ages 12-17) associated with the sampled adult[2], one child and one adolescent were randomly sampled, so up to three interviews could have been completed in each sampled household. The sampled adult was interviewed, and the parent or guardian most knowledgeable about the health and care of the sampled child was interviewed. The sampled adolescent responded for him or herself, but only after a parent or guardian gave permission for the interview. Table 1-2 shows the number of completed adult, child, and adolescent interviews in CHIS 2003, by the type of sample (RDD or supplemental sample).

[2] Only children for whom the sampled adult was parent or legal guardian were sampled. The CHIS 2003 sample weights account for this sampling procedure.

Table 1-2. Number of completed CHIS 2003 interviews by type of sample, instrument

| Type of sample | Adult | Child | Adolescent |
|---|---|---|---|
| Total RDD + supplemental cases | 42,044 | 8,526 | 4,010 |
| RDD | 41,818 | 8,480 | 3,996 |
| Supplemental samples: | | | |
| Korean | 112 | 24 | 6 |
| Vietnamese | 114 | 22 | 8 |

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

Interviews done in all languages were administered using Westat's computer-assisted telephone interviewing (CATI) system. The average adult interview took 33 minutes to complete. The average child and adolescent interviews took 14 minutes and 21 minutes, respectively. Interviews in the non-English languages generally took longer to complete. Approximately 11 percent of the adult interviews were completed in a language other than English, as were 21 percent of all child (parent proxy) interviews and 7 percent of all adolescent interviews.

Table 1-3 shows the major topic areas for each of the three survey instruments (adult, child, and adolescent).

## 1.4 Response Rate

The overall response rate for CHIS 2003 is a composite of the screener completion rate (i.e., success in introducing the survey to a household and randomly selecting an adult to be interviewed), and the extended interview completion rate (i.e., success in getting the selected person to complete the full interview). To maximize the response rate, especially at the screener stage, an advance letter (in five languages) was mailed to all sampled telephone numbers for which an address could be obtained from reverse directory services. An advance letter was mailed for approximately 72 percent of the sampled telephone numbers. In 2003, the screener completion rate was 55.9 percent[3], and the rate was higher for those households that could be sent the advance letter. The extended interview completion rate was 60.0 percent for the adult survey. Multiplying the screener and extended rates gives an overall response rate of 33.5 percent. Response rates vary by sampling stratum.

[3] In CHIS 2003, households that refused at the screener level were subsampled and only the subsampled households were called again in an attempt to convert them to respondents. The response rates are weighted to account for this subsampling.

Table 1-3. CHIS 2003 Survey topic areas by instrument

| **HEALTH STATUS** | **ADULT** | **TEEN** | **CHILD** |
|---|---|---|---|
| General health status, height and weight | ✓ | ✓ | ✓ |
| Emotional health | | ✓ | |
| Days missed from school due to health problems | | ✓ | ✓ |
| **HEALTH CONDITIONS** | **ADULT** | **TEEN** | **CHILD** |
| Asthma | ✓ | ✓ | ✓ |
| Heart disease, high blood pressure, epilepsy | ✓ | | |
| Diabetes | ✓ | ✓ | |
| Physical disability/need for special equipment | ✓ | ✓ | ✓ |
| Elder health (stroke, falls, incontinence) | ✓ | | |
| Parental concerns with child development, attention deficit disorder (ADD) | | | ✓ |
| **HEALTH BEHAVIORS** | **ADULT** | **TEEN** | **CHILD** |
| Dietary intake | | ✓ | ✓ |
| Physical activity and exercise | | ✓ | ✓ |
| Walking for transportation and leisure | ✓ | | |
| File and pneumonia immunization | ✓ | | |
| Alcohol and tobacco use | ✓ | ✓ | |
| Drug use | | ✓ | |
| Sexual behavior, STD testing, birth control practices | ✓ | ✓ | |
| **WOMEN'S HEALTH** | **ADULT** | **TEEN** | **CHILD** |
| Pap test screening, mammography screening, self-breast exam | ✓ | | |
| Emergency contraception, pregnancy status | ✓ | ✓ | |
| Menopause, hormone replacement therapy (HRT) | ✓ | | |
| **CANCER HISTORY AND PREVENTION** | **ADULT** | **TEEN** | **CHILD** |
| Cancer history of respondent | ✓ | | |
| Colon cancer screening, prostrate cancer (PSA) test | ✓ | | |
| **DENTAL HEALTH** | **ADULT** | **TEEN** | **CHILD** |
| Last dental visit, could not afford care, missed school/work days | ✓ | ✓ | ✓ |
| Dental insurance coverage | ✓ | ✓ | ✓ |
| **INJURY/VIOLENCE** | **ADULT** | **TEEN** | **CHILD** |
| Serious injuries (frequency, cause) | | ✓ | ✓ |
| Injury prevention behaviors (bike helmets, seatbelts) | | ✓ | ✓ |
| Infant-toddler home safety | | | ✓ |
| Interpersonal violence | | ✓ | |

Table 1-3. (Continued)

| ACCESS TO AND USE OF HEALTH CARE | ADULT | TEEN | CHILD |
|---|---|---|---|
| Usual source of care, visits to medical doctor | ✓ | ✓ | ✓ |
| Emergency room visits | ✓ | ✓ | ✓ |
| Delays in getting care (prescriptions, tests, treatment) | ✓ | ✓ | ✓ |
| Health care discrimination due to race or ethnic group | ✓ | | |
| Communication problems with doctor | ✓ | ✓ | ✓ |
| Ability and parental knowledge of teen contacting a doctor | | ✓ | |
| Child immunization reminders | | | ✓ |
| **HEALTH INSURANCE** | **ADULT** | **TEEN** | **CHILD** |
| Current insurance coverage, spouse's coverage, who pays for it | ✓ | ✓ | ✓ |
| Health plan enrollment, characteristics and assessment of plan | ✓ | ✓ | ✓ |
| Whether employer offers coverage, respondent/spouse eligibility | ✓ | | |
| Coverage over past 12 months | ✓ | ✓ | ✓ |
| Reasons for lack of insurance | ✓ | ✓ | ✓ |
| **EMPLOYMENT** | **ADULT** | **TEEN** | **CHILD** |
| Employment status, spouse's employment status | ✓ | | |
| Work in last week, industry and occupation | ✓ | | |
| Hours worked at all jobs | ✓ | ✓ | |
| **INCOME** | **ADULT** | **TEEN** | **CHILD** |
| Respondent and spouse's earnings last month before taxes | ✓ | | |
| Household income (annual before taxes) | ✓ | | |
| Number of persons supported by household income | ✓ | | |
| Assets | ✓ | | |
| **PUBLIC PROGRAM ELIGIBILITY** | **ADULT** | **TEEN** | **CHILD** |
| Household poverty level (100%, 130%, 200%, 300% FPL) | ✓ | | |
| Program participation (TANF, CalWorks, Public Housing, Food Stamps, SSI, SSDI, WIC) | ✓ | ✓ | ✓ |
| Assets, alimony/child support/social security/pension | ✓ | | |
| Reason for Medi-Cal non-participation among potential eligibles | ✓ | ✓ | ✓ |
| **FOOD INSECURITY/HUNGER** | **ADULT** | **TEEN** | **CHILD** |
| Availability of food in household over past 12 months | ✓ | | |
| **PARENTAL INVOLVEMENT** | **ADULT** | **TEEN** | **CHILD** |
| Parental presence after school, parental knowledge of whereabouts and activities | | ✓ | |
| Child's activities with family | | | ✓ |
| **NEIGHBORHOOD AND HOUSING** | **ADULT** | **TEEN** | **CHILD** |
| Neighborhood cohesion | ✓ | | |
| Neighborhood safety | ✓ | ✓ | |
| Neighborhood characteristics for children | | | ✓ |
| Length of time at current address/neighborhood, type of housing | ✓ | | |
| Home ownership, number of rooms, amount of mortgage/rent | ✓ | | |

Table 1-3. (Continued)

| CHILD CARE | ADULT | TEEN | CHILD |
|---|---|---|---|
| Current child care arrangements | | | ✓ |
| Child care over past 12 months | | | ✓ |
| Reason for lack of childcare | | | ✓ |
| **RESPONDENT CHARACTERISTICS** | **ADULT** | **TEEN** | **CHILD** |
| Age, gender, height, weight, education | ✓ | ✓ | ✓ |
| Race and ethnicity | ✓ | ✓ | ✓ |
| Marital status | ✓ | | |
| Sexual orientation | ✓ | | |
| Citizenship, immigration status, country of birth, English language proficiency | ✓ | ✓ | ✓ |

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

The CHIS response rate is comparable to response rates of other scientific telephone surveys in California, such as the California Behavioral Risk Factor Surveillance System (BRFSS) survey. California as a whole, and the state's urban areas in particular, are among the most difficult parts of the nation in which to conduct telephone interviews. Survey response rates tend to be lower in California than nationally, and over the past decade response rates have been declining both nationally and in California.

One way to judge the representativeness of a population survey is to "benchmark" its results against those of other reliable data sources. The CHIS 2001 sample yielded unweighted and weighted population distributions and rates that are comparable to those obtained from other sources. The demographic characteristics of the CHIS 2001 sample (such as race, ethnicity, and income) are very similar to those obtained from 2000 Census data. CHIS 2001 respondents also have health characteristics and behaviors that also are very similar to those found in other reliable surveys, such as the California BRFSS. An extensive benchmarking project is being undertaken for the 2003 California Health Interview Survey.

Adults who had completed at least 80 percent of the questionnaire (i.e., through Section I on health insurance) after all followup attempts were exhausted to complete the full questionnaire were counted as "complete." At least some items in the employment and income series or public program eligibility and food insecurity series are missing from these cases.

Proxy interviews were allowed for frail and ill persons over the age of 65 to avoid biases for health estimates for elderly persons that might otherwise result. Eligible selected persons were

recontacted and offered a proxy option. For 171 elderly adults, a proxy interview was completed by either a spouse/partner or adult child. Only a subset of questions identified as appropriate for a proxy respondent were administered. (Note: The questions not administered are identified in their response set as being skipped (denoted by a value of "-2") because a proxy is responding for the selected person.)

## 1.5 Weighting the Sample

To produce population estimates for the RDD CHIS results, weights are applied to the sample data to compensate for a variety of factors, some directly resulting from the design and administration of the survey. The sample is weighted to represent the non-institutionalized population for each sampling stratum and statewide. Sample weighting was carried out in CHIS 2003 to accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons (Note: telephone numbers for which addresses could be found and advance letters mailed were assigned a higher probability of selection than those without addresses);
- Reduce biases occurring because nonrespondents may have different characteristics than respondents;
- Adjust, to the extent possible, for undercoverage in the sampling frames and in the conduct of the survey; and
- Reduce the variance of the estimates by using auxiliary information.

As part of the weighting process, a household weight was created for all households that completed the screener interview. This household weight is the product of the "base weight" or the inverse of the probability of selection of the telephone number and adjustment factors computed for the following weight adjustments:

- Subsampling for numbers with addresses;
- Multiple chances of being selected in the RDD and supplemental samples;
- Unknown residential status;
- Subsampling screener refusals for conversion attempt;
- Screener interview nonresponse;

- Multiple telephone numbers; and
- Household poststratification.

The resulting poststratified household weight was used to compute a person-level weight. This person-level weight includes weight adjustments for the within-household sampling of persons and nonresponse. The final step is to adjust the person-level weight using a raking method so that the CHIS estimates are consistent with population control totals. Raking is an iterative procedure that forces the CHIS weights to sum to known totals from auxiliary data sources. The procedure requires iteration to make sure all the control totals or dimensions of raking are simultaneously satisfied (within a specified tolerance).

The control totals or raking dimensions used in CHIS 2003 were created primarily from the 2003 California Department of Finance estimates of the numbers of persons by age, race, and sex, and from the 2000 Census of Population counts from the U.S. Census Bureau. The 14 dimensions are combinations of demographic variables (age, sex, race, and ethnicity), geographic variables (county, city, and, in Los Angeles County, Service Planning Area), household composition (presence of children and adolescents in the household), and socio-economic variables (home ownership and education). The socio-economic variables are included to reduce biases associated with excluding households without a telephone number from the survey. One of the limitations of using the Department of Finance data is that it includes about 2.4 percent of the population of California who live in "group quarters" (i.e., persons living with 9 or more unrelated persons). These persons were excluded from the CHIS sample and, as a result, the number of persons living in group quarters had to be estimated and removed from the control totals prior to raking.

## 1.6 Imputation Methods

To enhance the utility of the CHIS 2003 data files, missing values were replaced through imputation for nearly every variable. This was a massive task designed to eliminate missing values in all source variables. Westat imputed values for variables used in the weighting process, and the UCLA staff imputed values where missing due to item nonresponse for nearly all other variables.

Two different imputation procedures were used by Westat prior to delivering the data to UCLA to fill in missing responses for items in CHIS 2003 that were essential for weighting the data. The

first imputation technique is a completely random selection from the observed distribution of the respondents. This method is used only for a few items when the percentage of the items that are missing is very small. For example, when imputing the missing values for self-reported age which had a very low item non-response rate, the distributions of the responses for age by type of interview (adult, child, or adolescent) were used to randomly assign an age using probabilities associated with these distributions.

The second technique is hot deck imputation without replacement. The hot deck approach is probably the most commonly used method for assigning values for missing responses in large-scale household surveys. With a hot deck, a value reported by a respondent for a particular item is assigned or donated to a "similar" person who did not respond to that item. The characteristics defining "similar" vary for different variables. To carry out hot deck imputation, the respondents to an item form a pool of donors, while the nonrespondents are a group of recipients. A recipient is matched to the subset pool of donors based on household and individual characteristics. A value for the recipient is then randomly imputed from one of the donors in the pool. Once a donor is used, it is removed from the pool of donors for that variable. Hot deck imputation was used to impute race, ethnicity, home ownership, and education in CHIS 2003.

The UCLA staff imputed missing values through a hierarchical sequential hot deck method with donor replacement. This method rank-orders the control variables from the most essential to the least essential, allowing the control variables to be dropped if the imputation conditions (such as minimal number of donors or no missingness in control variables) are not met in the imputation process. The control variables are dropped one at a time sequentially, starting from the least essential. CHIS incorporated an automated data quality control check both before and after the imputation process.

Imputation flags for CHIS source variables are included in separate data files to identify all imputed values.

## 1.7 Methodology Report Series

A series of five methodology reports are available with more detail about the methods used in CHIS 2003:

- Report 1 – Sample Design;
- Report 2 – Data Collection Methods;
- Report 3 – Data Processing Procedures;
- Report 4 – Response Rates; and
- Report 5 – Weighting and Variance Estimation.

For further information on CHIS data and the methods used in the survey, visit the California Health Interview Survey Web site at www.CHIS.ucla.edu or contact CHIS at CHIS@ucla.edu.

## 2. DATA EDITING PROCEDURES

Survey data for CHIS 2003 were collected using a computer-assisted telephone interview (CATI) system. In a CATI environment, the data collection and interview process is controlled using a series of computer programs designed to ensure consistency and quality. (*CHIS 2003 Methodology Series: Report 2 - Data Collection Methods* provides a thorough discussion of the interview process and a description of how the survey data were collected.) The CATI system programming determines which questions are asked based on household composition, respondent characteristics or preceding answers, and the order in which the questions are presented to interviewers. The system also presents the response options that are available for recording respondents' answers.

CATI range and logic edits do much to help ensure the integrity of the data during collection. This editing at the time of the interview greatly reduces the need to recontact respondents to verify responses and allows questionable entries to be reviewed in real time with the respondent as part of the collection process. Although the CATI system virtually eliminates out-of-range responses and many other anomalies, some consistency and edit issues may arise. For example, interviewers may note concerns or problems that must be handled by data preparation staff after the interview is complete. Updating activities require that both manual and machine editing procedures be developed to correct interviewer, respondent, and CATI program errors and to check that updates made by data preparation staff were input correctly. Because data editing resulted in changes to the survey data, specific quality control procedures were implemented. CHIS 2003 survey data were carefully examined and edited before delivering final data files. Quality control procedures involved limiting the number of staff who made updates, using the CATI specifications to resolve issues in complex questionnaire sections, carefully checking updates, and performing computer runs to identify inconsistencies or illogical patterns in the data.

The data editing procedures for CHIS 2003 consisted of four main tasks: (1) managing and resolving problem cases, (2) reading and using interviewer comments to make data updates, (3) coding questions with text strings (other specify responses), and (4) verifying data editing updates. The final step was to convert the edited data from the CATI system to the SAS data delivery files. The sections below describe each of these processes.

## 2.1 Resolving Problem Cases

One of the important tasks for ensuring high-quality data was managing and resolving problem cases. The data preparation staff, as well as project staff and staff from the Telephone Research Centers (TRCs), worked collectively to resolve problem cases. In this section, the method interviewers used to communicate problems is described, along with the system used by data editing and preparation staff to update or modify the data.

An interviewer who experienced a problem while working a case during data collection could alert the project team in one of two ways. One method was to fill out a hard copy problem sheet for the case. Problem sheets from all the TRCs were sent to a single staff member who distributed them to the appropriate department or project staff person. Data preparation staff often used these problem sheets as a guide to review cases and to make certain that any required updates were made accurately.

The second method of communicating problems was to assign a specific result code to cases within the CATI system, obviating the need for a hard copy problem sheet. The problem result code category had three sub-categories for special queues to which these problem cases could be assigned for review. These sub-categories were used to indicate the person responsible for investigating the case further—TRC staff, project staff, or data processing staff. Problem cases were reviewed electronically by a TRC supervisor and either re-fielded to the interviewing staff or distributed to the appropriate TRC, data processing, or project staff.

Database updates were unnecessary for some problems and these cases could simply be released for general interviewing accompanied by an appropriate message. If, for example, an adult extended interview was stopped during the middle of Section D, the interviewer would enter a detailed comment explaining why the case could not proceed (e.g., "Respondent wanted to change several answers. I was unable to backup properly"). The solution for these types of cases was to re-field the interview with a message stating, "Case will restart in Section D. Re-ask beginning with screen AE15." (Note that questions in CHIS 2001 that were also asked in CHIS 2003 retained their original CATI screen numbering. In this example, the first question in Section D for CHIS 2003 is screen AE15 from CHIS 2001.) Most restart cases were made available to the general interviewing staff. For unusual or complex problems, the case could be assigned to a specific interviewer with experience in handling these types of problems.

Some examples of cases reviewed by project staff were those in which an error was made in enumerating a household member or when the person named as most knowledgeable about the sampled child needed to be changed. Other types of problems required special interviewer handling, even after changes were made to the database.

One specific category of problems—enumeration errors where some household members were either incorrectly identified or their characteristics were entered in error—was somewhat more challenging than other types of errors to resolve. These problems touched upon sampling issues and, therefore, required careful review and treatment in order to preserve the integrity of the study's sampling procedures. These problem cases were resolved by reviewing the specific issues and case details with a project sampling statistician and making the appropriate changes to the data or by re-fielding the case. It is important to note, however, that very few cases had enumeration errors.

## 2.2 Interviewer Comments

Another important data editing task was reading and using interviewer comments. Comments are text phrases typed by interviewers in special entry windows in CATI when the respondent makes a statement that the interviewer wants to record but is unable to enter as a standard response in the instrument. For CHIS 2003, sometimes these phrases were merely an elaboration of a previously recorded response, an expression of opinion, or comments unrelated to the survey. Interviewer comments did not necessarily require modifying or updating survey responses. Other times, comments were substantive to data quality and indicated that an update was needed.

Comments were also used to identify specific responses that could not be coded using the existing response option set. Although project staff resolved the vast majority of these situations, several out-of-range and unanticipated responses required clarification. These issues were discussed and resolved by the Westat and UCLA CHIS project teams. Several items from CHIS 2001 that initially elicited unexpected responses were administered again in 2003. Based on the 2001 survey effort, the response option set for some items was amended or updated before fielding them for CHIS 2003. For example, question AG15 (item AH47 from CHIS 2001) in the adult extended interview asked; "What is the highest grade of education you have completed and received credit for?" Response code "30 – no formal education" was added to the list of categories displayed to interviewers. Updates to the response options helped to reduce the number of interviewer comments and lessen the amount of data preparation work.

The number and the extent of unanticipated responses were somewhat less for CHIS 2003 than in 2001. For CHIS 2003, data preparation staff were better prepared to code responses from interviewer comments and from "other specify" (see section 2.3 Coding with Test Strings) answers given their experience in handling similar or identical situations that occurred during CHIS 2001.

Additional response options/values were added to a limited number of survey items in CHIS 2003. For example, question AH13 from the adult extended interview asked respondents about their usual source of health care; "What kind of place do you go to most often – a doctor's office, a clinic or hospital clinic, an emergency room, or some other place?" The initial list of response options did not include a category for those who sought care in a non-traditional setting. The response option "alternative medical care" was subsequently added to the instrument.

A question that elicited several responses outside the standard response set in both CHIS 2001 and CHIS 2003 was AB25 (item AB34 from CHIS 2001) from the adult extended interview, "Has a doctor ever told you that you have any kind of heart disease?" Interviewers recorded responses in the comment field for this item such as "I have an irregular heart beat," "my doctor said I have a heart murmur," and "heart arrhythmia." Westat project and data preparation staff consulted with UCLA to accurately code such responses, and the interviewing staff subsequently received instructions for coding these responses.

Weekly meetings between data preparation and project staff were conducted during data collection to discuss data-related issues, review comments, and to establish case-specific procedures for handling pending or interim problem cases. Comments and cases under review included both completed and incomplete (interim status) interviews. Lists of out-of-range responses and responses that Westat staff could not unambiguously code were forwarded to the UCLA CHIS staff for review. The listings included suggestions and recommendations for new response options and for potential wording changes in an effort to clarify some survey items.

## 2.3 Coding with Text Strings

Most items in the CHIS 2003 survey had only closed-ended response options, so coding of open-ended responses was not needed. The survey had a number of other-specify type questions,

however, that required coding of narrative text strings recorded by interviewers. Other-specify questions were those in which a question had specific response categories but also allowed for text or values to be typed into an "other" category. For example, question AA5 in the adult extended interview asked respondents "And what is your Latino or Hispanic ancestry or origin? Such as Mexican, Salvadoran, Cuban, Honduran -- and if you have more than one, tell me all of them." An "other" category was available for responses that fell outside the list of categories that were read as a part of the question. Additional questions with an "other" category from the CHIS 2003 adult extended interview included;

ethnic ancestry questions (AA5, AA5E, AA5E1),
tribal names (AA5B, AA5D),
flu shot, where shot received (AB57, AB58),
how breast cancer was found (AB60),
sexual orientation (AD46),
STDs tested for (AD48),
country of birth (AH33, AH34, AH35),
languages spoken at home (AH36),
place visited for health care (AH3),
health insurance coverage items (AI15, AI15A, AI17A, AI45, AI45A, AI24, AI36), and
child/adolescent health insurance coverage items (CF7, IA7, CF18, IA18, CF29, IA29).

Westat data preparation staff reviewed these responses and up-coded them to the existing categories whenever possible. A limited number of survey items were augmented with additional response codes to accommodate answers recorded in the other-specify category and from comments. The updated response codes for these items are given in Table 2-1.

With the exception of the industry and occupation items in the adult extended interview, CHIS 2003 did not collect open-ended responses that required a specially developed coding scheme or structure. Many survey items, however, collected amounts or values such as the respondent's age, weight at age 18, etc. For such items, the CATI system utilized "soft-" and "hard-range" edit specifications.

Table 2-1. Response codes added to CHIS 2003

| Questionnaire version | Variable Name | Question Name | New code | Response description |
|---|---|---|---|---|
| Adult | AB57 | QA03_48 | 7 | Place of worship |
| Adult | AB58 | QA03_49 | 9 | Scared, afraid, do not like flu shots |
| | | | 10 | No reason, never get them, don't want |
| | | | 11 | Will get shot soon |
| Adult | AD48 | QA03_95 | 7 | All STDs, all of them, complete panel |
| Adult | AH3 | QA03_171 | 6 | Alternative medical care |
| Child | CH13 | QC03_143 | 3 | Mother deceased |
| | | | 4 | Mother does not live in US |
| Child | CH16 | QC03_147 | 3 | Father deceased |
| | | | 4 | Father does not live in US |

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

CATI edit specifications were initially prepared by Westat staff and then forwarded to UCLA for review, comment, and approval. The specifications were then implemented to improve data quality by informing interviewers when an out-of-acceptable-range or seemingly improbable response was recorded. Edit specifications enabled interviewers to identify and correct potential errors with the respondent during the interview and eliminated the need for a call back.

Soft-range edits were activated during the interview when the respondent gave an unlikely response (a value outside the specified range). The CATI system responded by placing a message on the screen and required the interviewer to re-enter the response. This system feature gives the interviewer an opportunity to verify that the response is recorded accurately, or, as needed, re-ask the question to be certain the respondent understood what was being asked. Hard-range edits prevented recording unacceptable values. For example, for a question on how many rooms were in the adult respondent's home, the soft range is 1-15, the hard range 1-99. During data collection, a small number of soft- and hard-range edit specifications were revised to accept the actual range of responses being collected.

In circumstances when the respondent insisted on giving a response that violated the soft- or hard-edit specifications, interviewers recorded the respondent's answer in the comment field and data preparation staff reviewed and updated the case as needed.

## 2.4 Verifying Data Updates

Updates to the original interview data were required due to a variety of circumstances as described above. Generally speaking, data updates and corrections were made to account for these situations including interviewer and respondent error, information captured in comments and "other-specify" fields, and problem sheets so that the final survey data reflected the most accurate information possible.

A series of techniques verified that survey updates were made accurately. First, the intended updates were recorded on a hard-copy printout or on an associated problem sheet. The CATI case identification number was also recorded to ensure that updates were associated with the appropriate case. This printout was checked for accuracy and for logical effects on any other questions or skip patterns in the questionnaire. Next, the updates were entered into the computer and verified again – matching the resulting information against the print-out. For more complicated circumstances, the data preparation staff carefully reviewed interviewer comments, messages, and problem descriptions to verify data updates.

An entry in an electronic transaction journal was created for each data update. Transaction journal entries maintained information such as the CATI case identification number, initial data value(s), the updated value(s), and the date that the update was made. The editing and verification process was performed throughout the data collection period and approximately 18,000 database values were updated and verified for CHIS 2003.

Cases with similar problems were reviewed together and then updated at one time in manageable batches. This process ensured consistency in the handling of discrete data problems. Following the series of updates, a program was run to check for the full set of errors that had been identified to date to ensure that data editing had not created any new errors. Frequency distributions and cross-tabulations of survey variables were used extensively by data preparation staff to verify data updates.

Structural edits designed to assess the integrity of the CATI database (i.e., verifying that all database records that should exist actually do exist, and those that should not exist do not), and, as necessary, edits that evaluated complex skip patterns were run periodically during data collection.

When discrepancies were discovered, problem cases were identified and reviewed, and updates were made as necessary. If data were incorrectly keyed in the database, the audit trail for the interview (a keystroke-by-keystroke record of all responses entered during the CATI interview) could be retrieved to determine the appropriate response. The interview audit trail is especially useful for reconstructing interviews that are interrupted unexpectedly by a power failure or system crash.

## 2.5 Data Conversion and Delivery

The final survey data were delivered to UCLA formatted as SAS data sets. The SAS data sets were created by converting the CATI database using a series of SAS macro programs. Initially, the CATI survey data are stored in a hierarchical database to improve data efficiency and enhance performance while interviewing. This conversion was accomplished using Westat's CATISAS macro program that extracts information stored in the CATI data dictionary (e.g., variable names, variable labels, allowable values, and formats) and then converts each of the CATI database segments into a "flat" SAS data set. Using the CATI data dictionary to define the SAS data set variables is advantageous because variables are stored in questionnaire order, allowing for meaningful presentation of the variables in frequency output and file listings without additional programming. SAS data sets created by the macro were later combined to facilitate processing and file delivery. After the survey data were converted from the CATI system, all further processing relied on the SAS system.

During the conversion process from CATI database elements to SAS files, edit checks were run on the entire database for diagnostic purposes. Frequencies for categorical data were also run and examined. These reviews were made to ensure that errors had not been inadvertently introduced to the data (i.e., no data were lost, no unexpected shifts in variable distributions occurred). In going from the CATI to SAS file organization, for example, frequency runs from the CATI database and the post-CATI SAS files were compared.

Data deliveries made to UCLA by Westat are summarized below.

- Screener/household membership variables;
- Questionnaire variables;

- Administrative/derived variables, such as classification, counter, and composite variables; and North American Industry Classification System/Standard Occupational Code;
- Weights (final sample weight and replicate weights); and
- Imputation flags.

The geocoded data were separately delivered to UCLA by Mapping Analytics as described in the next chapter, Geographic and Industry and Occupation Coding.

# 3. GEOGRAPHIC AND INDUSTRY AND OCCUPATION CODING

For CHIS 2003, Westat was responsible for delivering coded survey data for items from the adult extended interview related to industry and occupation (coded by the Census Bureau) and to geographic location of residence (coded by Mapping Analytics). This report section summarizes these two coding processes.

## 3.1 Geographic Coding

The CHIS 2003 adult extended interview asked all respondents the name of the county where they lived: "To be sure we are covering the entire state, what county do you live in?" (AH42). In addition, for cases in which an address had been matched to the sampled telephone number[4], interviewers verified the street address and Zip Code with the adult respondent (AO1) and then collected the name of a nearby cross-street (AM9).

If there was no matched address for a given case, respondents were asked to provide their Zip Code (AM7), their street address (AO2) and then the name of a nearby cross-street (AM9). Adult respondents who refused to provide a complete street address with house number were asked just for the name of the street they lived on (AM8) and the nearest cross street. Additional information about the CHIS 2003 geocoding of households is available (please see *Methodology Brief – 2003 Geocoding of Households* at http://www.chis.ucla.edu/methods.html ).

Because telephone numbers were assigned to sampling strata (see *CHIS 2003 Methodology Series: Report 1 - Sample Design*) based on the telephone area code and exchange, and some exchanges serve more than one county or city, the actual stratum where the respondent resides may differ from the sampling stratum. Both to monitor the sample yield during data collection and to ensure that the analysis file reflects the sampled person's actual residence, it was important to assign each adult who completed the extended survey to the correct self-reported stratum.

[4] The verification was not done if the telephone number was unlisted or if the sample vendor indicated that the number was on the "do not call" list.

Two questions from the adult extended interview were used to make the stratum assignment during data collection. The two survey questions are:

AH42. To be sure we are covering the entire state, what county do you live in?"

and

AM7. "What is your Zip Code?"

The self-reported stratum for monitoring sample yield was derived from the self-reported county values (AH42) for all strata except the cities of Oakland and Hayward in Alameda County. For these city strata, the self-reported stratum assignment is based on self-reported city and Zip Code. Table 3-1 shows a listing of Zip Codes within each stratum.

The final self-reported stratum for the analysis file was determined by applying the geocodes developed by Mapping Analytics with input from UCLA CHIS staff. See *CHIS 2003 Methodology Series: Report 5 - Weighting and Variance Estimation*, Section 8.5, for a fuller discussion of this process.

The final distribution of adult extended completed cases by self-reported and original sampling stratum is presented in Table 3-2. Generally, the frequency counts show that there is good correspondence between the original sampling stratum and the self-reported stratum with the exception of the three strata from Alameda County (7.1 Hayward, 7.2 Oakland, 7.3 remainder of Alameda). In these three areas, the sampled telephone exchanges were geographically distributed more widely than anticipated and many respondents reported living in an area outside of where they were sampled. The self-reported stratum may differ from the original sampling stratum, however, because the sampling stratum may have been incorrect or the respondent may have incorrectly reported the county of residence.

Table 3-1. Zip Codes within sampling stratum

| Stratum | | Zip Code |
|---|---|---|
| 1.0 | Los Angeles | 90001, 90002, 90003, 90004, 90005, 90006, 90007, 90008, 90009, 90010, 90011, 90012, 90013, 90014, 90015, 90016, 90017, 90018, 90019, 90020, 90021, 90022, 90023, 90024, 90025, 90026, 90027, 90028, 90029, 90030, 90031, 90032, 90033, 90034, 90035, 90036, 90037, 90038, 90039, 90040, 90041, 90042, 90043, 90044, 90045, 90046, 90047, 90048, 90049, 90050, 90051, 90052, 90053, 90054, 90055, 90056, 90057, 90058, 90059, 90060, 90061, 90062, 90063, 90064, 90065, 90066, 90067, 90068, 90069, 90070, 90071, 90072, 90073, 90074, 90075, 90076, 90077, 90078, 90079, 90080, 90081, 90082, 90083, 90084, 90086, 90087, 90088, 90089, 90091, 90093, 90094, 90095, 90096, 90097, 90099, 90101, 90102, 90103, 90174, 90185, 90189, 90201, 90202, 90209, 90210, 90211, 90212, 90213, 90220, 90221, 90222, 90223, 90224, 90230, 90231, 90232, 90233, 90239, 90240, 90241, 90242, 90245, 90247, 90248, 90249, 90250, 90251, 90254, 90255, 90260, 90261, 90262, 90263, 90264, 90265, 90266, 90267, 90270, 90272, 90274, 90275, 90277, 90278, 90280, 90290, 90291, 90292, 90293, 90294, 90295, 90296, 90301, 90302, 90303, 90304, 90305, 90306, 90307, 90308, 90309, 90310, 90311, 90312, 90313, 90397, 90398, 90401, 90402, 90403, 90404, 90405, 90406, 90407, 90408, 90409, 90410, 90411, 90501, 90502, 90503, 90504, 90505, 90506, 90507, 90508, 90509, 90510, 90601, 90602, 90603, 90604, 90605, 90606, 90607, 90608, 90609, 90610, 90612, 90623, 90630, 90631, 90637, 90638, 90639, 90640, 90650, 90651, 90652, 90659, 90660, 90661, 90662, 90665, 90670, 90671, 90701, 90702, 90703, 90704, 90706, 90707, 90710, 90711, 90712, 90713, 90714, 90715, 90716, 90717, 90723, 90731, 90732, 90733, 90734, 90744, 90745, 90746, 90747, 90748, 90749, 90755, 90801, 90802, 90803, 90804, 90805, 90806, 90807, 90808, 90810, 90813, 90814, 90815, 90822, 90831, 90832, 90833, 90834, 90835, 90840, 90842, 90844, 90845, 90846, 90847, 90848, 90853, 90888, 90895, 90899, 91001, 91003, 91006, 91007, 91009, 91010, 91011, 91012, 91016, 91017, 91020, 91021, 91023, 91024, 91025, 91030, 91031, 91040, 91041, 91042, 91043, 91046, 91066, 91077, 91101, 91102, 91103, 91104, 91105, 91106, 91107, 91108, 91109, 91110, 91114, 91115, 91116, 91117, 91118, 91121, 91123, 91124, 91125, 91126, 91129, 91131, 91175, 91182, 91184, 91185, 91186, 91187, 91188, 91189, 91191, 91201, 91202, 91203, 91204, 91205, 91206, 91207, 91208, 91209, 91210, 91214, 91221, 91222, 91224, 91225, 91226, 91301, 91302, 91303, 91304, 91305, 91306, 91307, 91308, 91309, 91310, 91311, 91312, 91313, 91316, 91321, 91322, 91324, 91325, 91326, 91327, 91328, 91329, 91330, 91331, 91333, 91334, 91335, 91337, 91340, 91341, 91342, 91343, 91344, 91345, 91346, 91350, 91351, 91352, 91353, 91354, 91355, 91356, 91357, 91361, 91362, 91363, 91364, 91365, 91367, 91371, 91372, 91376, 91380, 91381, 91382, 91383, 91384, 91385, 91386, 91387, 91388, 91390, 91392, 91393, 91394, 91395, 91396, 91399, 91401, 91402, 91403, 91404, 91405, 91406, 91407, 91408, 91409, 91410, 91411, 91412, 91413, 91416, 91423, 91426, 91436, 91470, 91482, 91495, 91496, 91497, 91499, 91501, 91502, 91503, 91504, 91505, 91506, 91507, 91508, 91510, 91521, 91522, 91523, 91526, 91601, 91602, 91603, 91604, 91605, 91606, 91607, 91608, 91609, 91610, 91611, 91612, 91614, 91615, 91616, 91617, 91618, 91702, 91706, 91709, 91711, 91714, 91715, 91716, 91722, 91723, 91724, 91731, 91732, 91733, 91734, 91735, 91740, 91741, 91744, 91745, 91746, 91747, 91748, 91749, 91750, 91754, 91755, 91756, 91759, 91765, 91766, 91767, 91768, 91769, 91770, 91771, 91772, 91773, 91775, 91776, 91778, 91780, 91788, 91789, 91790, 91791, 91792, 91793, 91795, 91797, 91799, 91801, 91802, 91803, 91804, 91841, 91896, 91899, 93243, 93510, 93532, 93534, 93535, 93536, 93539, 93543, 93544, 93550, 93551, 93552, 93553, 93560, 93563, 93584, 93586, 93590, 93591, 93599, 96056 |

Table 3-1. Zip Codes within sampling stratum (continued)

| Stratum | | Zip Code |
|---|---|---|
| 2 | San Diego | 91901, 91902, 91903, 91905, 91906, 91908, 91909, 91910, 91911, 91912, 91913, 91914, 91915, 91916, 91917, 91921, 91931, 91932, 91933, 91934, 91935, 91941, 91942, 91943, 91944, 91945, 91946, 91947, 91948, 91950, 91951, 91962, 91963, 91976, 91977, 91978, 91979, 91980, 91987, 91990, 92003, 92004, 92007, 92008, 92009, 92010, 92011, 92013, 92014, 92018, 92019, 92020, 92021, 92022, 92023, 92024, 92025, 92026, 92027, 92028, 92029, 92030, 92033, 92036, 92037, 92038, 92039, 92040, 92046, 92049, 92051, 92052, 92054, 92055, 92056, 92057, 92058, 92059, 92060, 92061, 92064, 92065, 92066, 92067, 92068, 92069, 92070, 92071, 92072, 92074, 92075, 92078, 92079, 92081, 92082, 92083, 92084, 92085, 92086, 92088, 92090, 92091, 92092, 92093, 92096, 92101, 92102, 92103, 92104, 92105, 92106, 92107, 92108, 92109, 92110, 92111, 92112, 92113, 92114, 92115, 92116, 92117, 92118, 92119, 92120, 92121, 92122, 92123, 92124, 92126, 92127, 92128, 92129, 92130, 92131, 92132, 92133, 92134, 92135, 92136, 92137, 92138, 92139, 92140, 92142, 92143, 92145, 92147, 92149, 92150, 92152, 92153, 92154, 92155, 92158, 92159, 92160, 92161, 92162, 92163, 92164, 92165, 92166, 92167, 92168, 92169, 92170, 92171, 92172, 92173, 92174, 92175, 92176, 92177, 92178, 92179, 92182, 92184, 92186, 92187, 92190, 92191, 92192, 92193, 92194, 92195, 92196, 92197, 92198, 92199 |
| 3 | Orange | 90620, 90621, 90622, 90623, 90624, 90630, 90631, 90632, 90633, 90638, 90680, 90720, 90721, 90740, 90742, 90743, 92602, 92603, 92604, 92605, 92606, 92607, 92609, 92610, 92612, 92614, 92615, 92616, 92617, 92618, 92619, 92620, 92623, 92624, 92625, 92626, 92627, 92628, 92629, 92630, 92637, 92646, 92647, 92648, 92649, 92650, 92651, 92652, 92653, 92654, 92655, 92656, 92657, 92658, 92659, 92660, 92661, 92662, 92663, 92672, 92673, 92674, 92675, 92676, 92677, 92678, 92679, 92683, 92684, 92685, 92688, 92690, 92691, 92692, 92693, 92694, 92697, 92698, 92701, 92702, 92703, 92704, 92705, 92706, 92707, 92708, 92709, 92710, 92711, 92712, 92725, 92728, 92735, 92780, 92781, 92782, 92799, 92801, 92802, 92803, 92804, 92805, 92806, 92807, 92808, 92809, 92811, 92812, 92814, 92815, 92816, 92817, 92821, 92822, 92823, 92825, 92831, 92832, 92833, 92834, 92835, 92836, 92837, 92838, 92840, 92841, 92842, 92843, 92844, 92845, 92846, 92850, 92856, 92857, 92859, 92861, 92862, 92863, 92864, 92865, 92866, 92867, 92868, 92869, 92870, 92871, 92885, 92886, 92887, 92899 |
| 4 | Santa Clara | 94022, 94023, 94024, 94035, 94039, 94040, 94041, 94042, 94043, 94085, 94086, 94087, 94088, 94089, 94090, 94301, 94302, 94303, 94304, 94305, 94306, 94309, 94310, 94550, 95002, 95008, 95009, 95011, 95013, 95014, 95015, 95020, 95021, 95023, 95026, 95030, 95031, 95032, 95033, 95035, 95036, 95037, 95038, 95042, 95044, 95046, 95050, 95051, 95052, 95053, 95054, 95055, 95056, 95070, 95071, 95076, 95101, 95102, 95103, 95106, 95108, 95109, 95110, 95111, 95112, 95113, 95114, 95115, 95116, 95117, 95118, 95119, 95120, 95121, 95122, 95123, 95124, 95125, 95126, 95127, 95128, 95129, 95130, 95131, 95132, 95133, 95134, 95135, 95136, 95137, 95138, 95139, 95140, 95141, 95142, 95148, 95150, 95151, 95152, 95153, 95154, 95155, 95156, 95157, 95158, 95159, 95160, 95161, 95164, 95170, 95171, 95172, 95173, 95190, 95191, 95192, 95193, 95194, 95196 |

Table 3-1. Zip Codes within sampling stratum (continued)

| Stratum | | Zip Code |
|---|---|---|
| 5 | San Bernardino | 91701, 91708, 91709, 91710, 91729, 91730, 91737, 91739, 91743, 91758, 91761, 91762, 91763, 91764, 91766, 91784, 91785, 91786, 91792, 91798, 92242, 92252, 92256, 92267, 92268, 92277, 92278, 92280, 92284, 92285, 92286, 92301, 92304, 92305, 92307, 92308, 92309, 92310, 92311, 92312, 92313, 92314, 92315, 92316, 92317, 92318, 92321, 92322, 92323, 92324, 92325, 92326, 92327, 92329, 92332, 92333, 92334, 92335, 92336, 92337, 92338, 92339, 92340, 92341, 92342, 92344, 92345, 92346, 92347, 92350, 92352, 92354, 92356, 92357, 92358, 92359, 92363, 92364, 92365, 92366, 92368, 92369, 92371, 92372, 92373, 92374, 92375, 92376, 92377, 92378, 92382, 92385, 92386, 92391, 92392, 92393, 92394, 92395, 92397, 92398, 92399, 92401, 92402, 92403, 92404, 92405, 92406, 92407, 92408, 92410, 92411, 92412, 92413, 92414, 92415, 92418, 92423, 92424, 92427, 92880, 93516, 93555, 93562, 93592 |
| 6 | Riverside | 91752, 92028, 92201, 92202, 92203, 92210, 92211, 92220, 92223, 92225, 92226, 92230, 92234, 92235, 92236, 92239, 92240, 92241, 92247, 92248, 92253, 92254, 92255, 92258, 92260, 92261, 92262, 92263, 92264, 92270, 92274, 92276, 92282, 92292, 92320, 92324, 92373, 92399, 92501, 92502, 92503, 92504, 92505, 92506, 92507, 92508, 92509, 92513, 92514, 92515, 92516, 92517, 92518, 92519, 92521, 92522, 92530, 92531, 92532, 92536, 92539, 92543, 92544, 92545, 92546, 92548, 92549, 92551, 92552, 92553, 92554, 92555, 92556, 92557, 92561, 92562, 92563, 92564, 92567, 92570, 92571, 92572, 92581, 92582, 92583, 92584, 92585, 92586, 92587, 92589, 92590, 92591, 92592, 92593, 92595, 92596, 92599, 92860, 92877, 92878, 92879, 92880, 92881, 92882, 92883 |
| 7.1 | Alameda (Hayward) | 94540, 94541, 94542, 94543, 94544, 94545, 94546, 94552, 94557 |
| 7.2 | Alameda (Oakland) | 94601, 94602, 94603, 94604, 94605, 94606, 94607, 94608, 94609, 94610, 94611, 94612, 94613, 94614, 94615, 94617, 94618, 94619, 94621, 94622, 94623, 94624, 94625, 94626, 94627, 94643, 94649, 94659, 94660, 94661, 94666 |
| 7.3 | Alameda (remainder) | 94501, 94502, 94514, 94536, 94537, 94538, 94539, 94550, 94551, 94555, 94560, 94566, 94568, 94577, 94578, 94579, 94580, 94586, 94587, 94588, 94620, 94662, 94701, 94702, 94703, 94704, 94705, 94706, 94707, 94708, 94709, 94710, 94712, 94720, 95304, 95377, 95391 |
| 8 | Sacramento | 94203, 94204, 94205, 94206, 94207, 94208, 94209, 94211, 94229, 94230, 94232, 94234, 94235, 94236, 94237, 94239, 94240, 94243, 94244, 94245, 94246, 94247, 94248, 94249, 94250, 94252, 94253, 94254, 94256, 94257, 94258, 94259, 94261, 94262, 94263, 94267, 94268, 94269, 94271, 94273, 94274, 94277, 94278, 94279, 94280, 94282, 94283, 94284, 94285, 94286, 94287, 94288, 94289, 94290, 94291, 94293, 94294, 94295, 94296, 94297, 94298, 94299, 94571, 95608, 95609, 95610, 95611, 95615, 95621, 95624, 95626, 95628, 95630, 95632, 95638, 95639, 95640, 95641, 95652, 95655, 95660, 95662, 95670, 95671, 95673, 95680, 95683, 95690, 95693, 95741, 95742, 95743, 95757, 95758, 95759, 95763, 95812, 95813, 95814, 95815, 95816, 95817, 95818, 95819, 95820, 95821, 95822, 95823, 95824, 95825, 95826, 95827, 95828, 95829, 95830, 95831, 95832, 95833, 95834, 95835, 95836, 95837, 95838, 95840, 95841, 95842, 95843, 95851, 95852, 95853, 95857, 95860, 95864, 95865, 95866, 95867, 95887, 95894, 95899 |

Table 3-1. Zip Codes within sampling stratum (continued)

| | Stratum | Zip Code |
|---|---|---|
| 9 | Contra Costa | 94506, 94507, 94509, 94511, 94513, 94514, 94516, 94517, 94518, 94519, 94520, 94521, 94522, 94523, 94524, 94525, 94526, 94527, 94528, 94529, 94530, 94531, 94547, 94548, 94549, 94551, 94553, 94556, 94561, 94563, 94564, 94565, 94569, 94570, 94572, 94575, 94582, 94583, 94595, 94596, 94597, 94598, 94706, 94707, 94708, 94801, 94802, 94803, 94804, 94805, 94806, 94807, 94808, 94820, 94850, 94875 |
| 10 | Fresno | 93210, 93234, 93242, 93245, 93313, 93602, 93605, 93606, 93607, 93608, 93609, 93611, 93612, 93613, 93616, 93618, 93619, 93620, 93621, 93622, 93624, 93625, 93626, 93627, 93628, 93630, 93631, 93634, 93640, 93641, 93642, 93646, 93648, 93649, 93650, 93651, 93652, 93654, 93656, 93657, 93660, 93662, 93664, 93667, 93668, 93675, 93701, 93702, 93703, 93704, 93705, 93706, 93707, 93708, 93709, 93710, 93711, 93712, 93714, 93715, 93716, 93717, 93718, 93720, 93721, 93722, 93724, 93725, 93726, 93727, 93728, 93729, 93740, 93741, 93744, 93745, 93747, 93750, 93755, 93760, 93761, 93764, 93765, 93771, 93772, 93773, 93774, 93775, 93776, 93777, 93778, 93779, 93780, 93784, 93786, 93790, 93791, 93792, 93793, 93794, 93844, 93888 |
| 11 | San Francisco | 94101, 94102, 94103, 94104, 94105, 94106, 94107, 94108, 94109, 94110, 94111, 94112, 94114, 94115, 94116, 94117, 94118, 94119, 94120, 94121, 94122, 94123, 94124, 94125, 94126, 94127, 94129, 94130, 94131, 94132, 94133, 94134, 94135, 94136, 94137, 94138, 94139, 94140, 94141, 94142, 94143, 94144, 94145, 94146, 94147, 94150, 94151, 94152, 94153, 94154, 94155, 94156, 94157, 94158, 94159, 94160, 94161, 94162, 94163, 94164, 94165, 94166, 94167, 94168, 94169, 94170, 94171, 94172, 94175, 94177, 94188, 94199 |
| 12 | Ventura | 90265, 91304, 91307, 91311, 91319, 91320, 91358, 91359, 91360, 91361, 91362, 91377, 93001, 93002, 93003, 93004, 93005, 93006, 93007, 93009, 93010, 93011, 93012, 93013, 93015, 93016, 93020, 93021, 93022, 93023, 93024, 93030, 93031, 93032, 93033, 93034, 93035, 93036, 93040, 93041, 93042, 93043, 93044, 93060, 93061, 93062, 93063, 93064, 93065, 93066, 93093, 93094, 93099, 93252 |
| 13 | San Mateo | 94002, 94003, 94005, 94010, 94011, 94012, 94013, 94014, 94015, 94016, 94017, 94018, 94019, 94020, 94021, 94025, 94026, 94027, 94028, 94029, 94030, 94031, 94037, 94038, 94044, 94045, 94059, 94060, 94061, 94062, 94063, 94064, 94065, 94066, 94067, 94070, 94071, 94074, 94080, 94083, 94096, 94098, 94099, 94128, 94303, 94307, 94308, 94401, 94402, 94403, 94404, 94405, 94406, 94407, 94408, 94409, 94497 |
| 14 | Kern | 93203, 93205, 93206, 93215, 93216, 93220, 93222, 93224, 93225, 93226, 93238, 93240, 93241, 93243, 93249, 93250, 93251, 93252, 93255, 93263, 93268, 93276, 93280, 93283, 93285, 93287, 93301, 93302, 93303, 93304, 93305, 93306, 93307, 93308, 93309, 93311, 93312, 93313, 93314, 93380, 93381, 93382, 93383, 93384, 93385, 93386, 93387, 93388, 93389, 93390, 93501, 93502, 93504, 93505, 93516, 93518, 93519, 93523, 93524, 93527, 93528, 93531, 93536, 93554, 93555, 93556, 93558, 93560, 93561, 93581, 93596, 96044 |
| 15 | San Joaquin | 94514, 95201, 95202, 95203, 95204, 95205, 95206, 95207, 95208, 95209, 95210, 95211, 95212, 95213, 95215, 95219, 95220, 95227, 95230, 95231, 95234, 95236, 95237, 95240, 95241, 95242, 95253, 95258, 95267, 95269, 95296, 95297, 95304, 95320, 95330, 95336, 95337, 95361, 95366, 95376, 95377, 95378, 95385, 95391, 95632, 95686, 95690 |

Table 3-1. Zip Codes within sampling stratum (continued)

| Stratum | | Zip Code |
|---|---|---|
| 16 | Sonoma | 94515, 94922, 94923, 94926, 94927, 94928, 94931, 94951, 94952, 94953, 94954, 94955, 94972, 94975, 94999, 95401, 95402, 95403, 95404, 95405, 95406, 95407, 95408, 95409, 95412, 95416, 95419, 95421, 95425, 95430, 95431, 95433, 95436, 95439, 95441, 95442, 95444, 95446, 95448, 95450, 95452, 95462, 95465, 95471, 95472, 95473, 95476, 95480, 95486, 95487, 95492, 95497 |
| 17 | Stanislaus | 95230, 95304, 95307, 95313, 95316, 95319, 95322, 95323, 95326, 95328, 95329, 95350, 95351, 95352, 95353, 95354, 95355, 95356, 95357, 95358, 95360, 95361, 95363, 95367, 95368, 95380, 95381, 95382, 95385, 95386, 95387, 95397 |
| 18 | Santa Barbara | 93013, 93014, 93067, 93101, 93102, 93103, 93105, 93106, 93107, 93108, 93109, 93110, 93111, 93116, 93117, 93118, 93120, 93121, 93130, 93140, 93150, 93160, 93190, 93199, 93252, 93254, 93427, 93429, 93434, 93436, 93437, 93438, 93440, 93441, 93454, 93455, 93456, 93457, 93458, 93460, 93463, 93464 |
| 19 | Solano | 94503, 94510, 94512, 94533, 94534, 94535, 94571, 94585, 94589, 94590, 94591, 94592, 95616, 95620, 95625, 95687, 95688, 95690, 95694, 95696 |
| 20 | Tulare | 93201, 93207, 93208, 93212, 93215, 93218, 93219, 93221, 93223, 93227, 93235, 93237, 93238, 93244, 93247, 93256, 93257, 93258, 93260, 93261, 93262, 93265, 93267, 93270, 93271, 93272, 93274, 93275, 93277, 93278, 93279, 93282, 93286, 93290, 93291, 93292, 93527, 93603, 93615, 93618, 93631, 93633, 93641, 93646, 93647, 93654, 93666, 93670, 93673 |
| 21 | Santa Cruz | 95001, 95003, 95005, 95006, 95007, 95010, 95017, 95018, 95019, 95033, 95041, 95060, 95061, 95062, 95063, 95064, 95065, 95066, 95067, 95073, 95076, 95077 |
| 22 | Marin | 94901, 94903, 94904, 94912, 94913, 94914, 94915, 94920, 94924, 94925, 94929, 94930, 94933, 94937, 94938, 94939, 94940, 94941, 94942, 94945, 94946, 94947, 94948, 94949, 94950, 94956, 94957, 94960, 94963, 94964, 94965, 94966, 94970, 94971, 94973, 94974, 94976, 94977, 94978, 94979, 94998 |
| 23 | San Luis Obispo | 93252, 93401, 93402, 93403, 93405, 93406, 93407, 93408, 93409, 93410, 93412, 93420, 93421, 93422, 93423, 93424, 93426, 93428, 93430, 93432, 93433, 93435, 93442, 93443, 93444, 93445, 93446, 93447, 93448, 93449, 93451, 93452, 93453, 93454, 93461, 93465, 93475, 93483 |
| 24 | Placer | 95602, 95603, 95604, 95626, 95631, 95648, 95650, 95658, 95661, 95663, 95668, 95677, 95678, 95681, 95692, 95701, 95703, 95713, 95714, 95715, 95717, 95722, 95736, 95746, 95747, 95765, 96140, 96141, 96143, 96145, 96146, 96148, 96161 |
| 25 | Merced | 93610, 93620, 93622, 93635, 93661, 93665, 95301, 95303, 95312, 95315, 95317, 95322, 95324, 95333, 95334, 95340, 95341, 95342, 95343, 95344, 95348, 95365, 95369, 95374, 95380, 93620, 93622, 93635, 93661, 93665, 95301, 95303, 95312, 95315, 95317, 95322, 95324, 95333, 95334, 95340, 95341, 95342, 95343, 95344, 95348, 95365, 95369, 95374, 95380, 95388 |
| 26 | Butte | 95901, 95914, 95916, 95917, 95925, 95926, 95927, 95928, 95929, 95930, 95938, 95940, 95941, 95942, 95948, 95954, 95958, 95965, 95966, 95967, 95968, 95969, 95973, 95974, 95976, 95978 |

Table 3-1. Zip Codes within sampling stratum (continued)

| Stratum | | Zip Code |
|---|---|---|
| 27 | Shasta | 96001, 96002, 96003, 96007, 96008, 96011, 96013, 96016, 96017, 96019, 96022, 96025, 96028, 96033, 96040, 96047, 96049, 96051, 96056, 96059, 96062, 96065, 96069, 96070, 96071, 96073, 96076, 96079, 96084, 96087, 96088, 96089, 96095, 96096, 96099 |
| 28 | Yolo | 95605, 95606, 95607, 95612, 95616, 95617, 95618, 95627, 95637, 95645, 95653, 95679, 95691, 95694, 95695, 95697, 95698, 95776, 95798, 95799 |
| 29 | El Dorado | 95613, 95614, 95619, 95623, 95629, 95633, 95634, 95635, 95636, 95651, 95656, 95664, 95667, 95672, 95682, 95684, 95709, 95720, 95721, 95726, 95735, 95762, 96142, 96150, 96151, 96152, 96154, 96155, 96156, 96157, 96158 |
| 30 | Imperial | 92004, 92222, 92225, 92227, 92231, 92232, 92233, 92243, 92244, 92249, 92250, 92251, 92257, 92259, 92266, 92273, 92274, 92275, 92281, 92283 |
| 31 | Napa | 94503, 94508, 94515, 94558, 94559, 94562, 94567, 94573, 94574, 94576, 94581, 94589, 94590, 94599, 95476 |
| 32 | Kings | 93202, 93204, 93212, 93230, 93232, 93239, 93242, 93245, 93246, 93266, 93631, 93656 |
| 33 | Madera | 93601, 93602, 93604, 93610, 93614, 93620, 93622, 93623, 93626, 93637, 93638, 93639, 93643, 93644, 93645, 93653, 93669, 93720 |
| 34 | Monterey, San Benito | 93426, 93450, 93451, 93901, 93902, 93905, 93906, 93907, 93908, 93912, 93915, 93920, 93921, 93922, 93923, 93924, 93925, 93926, 93927, 93928, 93930, 93932, 93933, 93940, 93942, 93943, 93944, 93950, 93953, 93954, 93955, 93960, 93962, 95004, 95012, 95039, 95076, 93210, 93930, 95004, 95023, 95024, 95043, 95045, 95075 |
| 35 | Del Norte, Humboldt | 95531, 95531, 95532, 95538, 95543, 95548, 95567, 95501, 95502, 95503, 95511, 95514, 95518, 95519, 95521, 95524, 95525, 95526, 95528, 95534, 95536, 95537, 95540, 95542, 95545, 95546, 95547, 95549, 95550, 95551, 95553, 95554, 95555, 95556, 95558, 95559, 95560, 95562, 95564, 95565, 95569, 95570, 95571, 95573, 95589 |
| 36 | Lassen, Modoc, Siskiyou, Trinity | 96006, 96009, 96056, 96068, 96109, 96113, 96114, 96117, 96119, 96121, 96123, 96127, 96128, 96130, 96132, 96136, 96137, 96006, 96015, 96054, 96056, 96101, 96104, 96108, 96110, 96112, 96115, 96116, 96134, 95568, 96014, 96023, 96025, 96027, 96031, 96032, 96034, 96037, 96038, 96039, 96044, 96050, 96057, 96058, 96064, 96067, 96085, 96086, 96091, 96094, 96097, 96134, 95526, 95527, 95543, 95552, 95563, 95595, 96010, 96024, 96041, 96046, 96048, 96052, 96076, 96091, 96093 |
| 37 | Lake, Mendocino | 95422, 95423, 95424, 95426, 95435, 95443, 95451, 95453, 95457, 95458, 95461, 95464, 95467, 95485, 95493, 95410, 95415, 95417, 95418, 95420, 95425, 95427, 95428, 95429, 95432, 95437, 95445, 95449, 95454, 95456, 95459, 95460, 95463, 95466, 95468, 95469, 95470, 95481, 95482, 95488, 95490, 95494, 95585, 95587, 95589 |
| 38 | Colusa, Glen, Tehama | 95912, 95932, 95939, 95950, 95955, 95957, 95970, 95979, 95987, 95913, 95920, 95939, 95943, 95951, 95963, 95970, 95988, 95963, 95973, 96021, 96022, 96029, 96035, 96055, 96059, 96061, 96063, 96074, 96075, 96076, 96078, 96080, 96090, 96092 |
| 39 | Sutter, Yuba | 95626, 95645, 95648, 95659, 95668, 95674, 95676, 95692, 95837, 95948, 95953, 95957, 95982, 95991, 95992, 95993, 95692, 95901, 95903, 95914, 95918, 95919, 95922, 95925, 95935, 95941, 95960, 95961, 95962, 95966, 95972, 95977, 95981 |

Table 3-1. Zip Codes within sampling stratum (continued)

| Stratum | | Zip Code |
|---|---|---|
| 40 | Plumas, Nevada, Sierra | 95602, 95712, 95724, 95728, 95924, 95945, 95946, 95949, 95959, 95960, 95975, 95977, 95986, 96111, 96160, 96161, 96162, 95915, 95923, 95934, 95947, 95956, 95971, 95980, 95981, 95983, 95984, 96020, 96103, 96105, 96106, 96122, 96129, 96135, 96137, 95910, 95922, 95936, 95944, 95960, 96105, 96118, 96124, 96125, 96126 |
| 41 | Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne | 95646, 96120, 95601, 95629, 95640, 95642, 95644, 95654, 95665, 95666, 95669, 95675, 95685, 95689, 95699, 95221, 95222, 95223, 95224, 95225, 95226, 95228, 95229, 95230, 95232, 95233, 95236, 95245, 95246, 95247, 95248, 95249, 95250, 95251, 95252, 95254, 95255, 95257, 92328, 92384, 92389, 93513, 93514, 93515, 93522, 93526, 93527, 93530, 93542, 93545, 93549, 93601, 93623, 93653, 95306, 95311, 95318, 95321, 95325, 95329, 95338, 95345, 95389, 93512, 93514, 93517, 93529, 93541, 93546, 96107, 96133, 95230, 95305, 95309, 95310, 95311, 95314, 95321, 95327, 95329, 95335, 95346, 95347, 95364, 95370, 95372, 95373, 95375, 95379, 95383 |

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

Table 3-2. Final distribution of adult extended completed cases by self-reported and original sampling stratum

| Stratum name | Sampling stratum count | Removed | Added | Final self-reported stratum count |
|---|---|---|---|---|
| 1 - Los Angeles | 10,350 | 61 | 74 | 10,363 |
| 2 - San Diego | 2,310 | 1 | 10 | 2,319 |
| 3 - Orange | 2,231 | 65 | 20 | 2,186 |
| 4 - Santa Clara | 1,340 | 4 | 59 | 1,395 |
| 5 - San Bernardino | 1,238 | 18 | 24 | 1,244 |
| 6 - Riverside | 1,180 | 5 | 11 | 1,186 |
| 7.1 - Hayward | 1,629 | 843 | 2 | 788 |
| 7.2 - Oakland | 1,975 | 211 | 89 | 1,853 |
| 7.3 - Remainder Alameda | 1,130 | 121 | 997 | 2,006 |
| 8 - Sacramento | 1,062 | 5 | 4 | 1,061 |
| 9 - Contra Costa | 820 | 4 | 81 | 897 |
| 10 - Fresno | 626 | 8 | 12 | 630 |
| 11 - San Francisco | 917 | 25 | 12 | 904 |
| 12 - Ventura | 617 | 10 | 23 | 630 |
| 13 - San Mateo | 609 | 43 | 30 | 596 |
| 14 - Kern | 537 | 0 | 12 | 549 |
| 15 - San Joaquin | 521 | 2 | 4 | 523 |
| 16 - Sonoma | 507 | 1 | 13 | 519 |
| 17 - Stanislaus | 549 | 22 | 4 | 531 |
| 18 - Santa Barbara | 504 | 9 | 2 | 497 |
| 19 - Solano | 510 | 14 | 7 | 503 |
| 20 - Tulare | 575 | 3 | 10 | 582 |
| 21 - Santa Cruz | 512 | 32 | 0 | 480 |
| 22 - Marin | 521 | 2 | 3 | 522 |
| 23 - San Luis Obispo | 503 | 4 | 7 | 506 |
| 24 - Placer | 507 | 10 | 16 | 513 |
| 25 - Merced | 520 | 7 | 24 | 537 |
| 26 - Butte | 564 | 4 | 7 | 567 |
| 27 - Shasta | 506 | 4 | 35 | 537 |
| 28 - Yolo | 517 | 8 | 5 | 514 |
| 29 - El Dorado | 503 | 4 | 7 | 506 |
| 30 - Imperial | 529 | 3 | 2 | 528 |
| 31 - Napa | 505 | 5 | 13 | 513 |
| 32 - Kings | 531 | 6 | 3 | 528 |
| 33 - Madera | 512 | 6 | 0 | 506 |
| 34. - Monterey, San Benito | 520 | 3 | 25 | 542 |
| 35. - Del Norte, Humboldt | 529 | 7 | 3 | 525 |
| 36. - Lassen, Modoc, Siskiyou, Trinity | 419 | 9 | 13 | 423 |
| 37. - Lake, Mendocino | 409 | 14 | 1 | 396 |
| 38. - Colusa, Glen, Tehama | 425 | 33 | 5 | 397 |
| 39. - Sutter, Yuba | 460 | 19 | 10 | 451 |
| 40. - Plumas, Nevada, Sierra | 403 | 19 | 6 | 390 |
| 41. - Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne | 412 | 11 | 0 | 401 |
| Total | 42,044 | 1,685 | 1,685 | 42,044 |

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

## 3.2 Industry and Occupation Coding

The adult extended interview for CHIS 2003 included three questions about occupation and industry. These same items were asked in CHIS 2001. The three questions are:

> AK4, "On your MAIN job, are you employed by: a private company, the government, OR are you self-employed, OR are you working without pay in a family business or farm?"
>
> AK5, "What kind of business or industry is this?"
>
> AK6, "What is the main kind of work YOU do?"

In CHIS 2001, respondents who reported working for the Federal, state, or local government in question AK4, were not asked the business or industry question (AK5). This was changed for CHIS 2003, and all respondents who indicated they were working were asked about their business or industry (AK5). All adult respondents who indicated that they were working were then asked about the main kind of work that they do (AK6).

The goal of the coding of occupation and industry was to produce codes that are consistent with other published industry and occupation coded surveys. Westat subcontracted with the Census Bureau for the coding effort to ensure consistency. Census Bureau staff coded the data provided by Westat using "Concepts and Methods used in Labor Force Statistics Derived from the Current Population Survey," a joint publication of the Bureau of Labor Statistics and the Bureau of the Census.

CHIS 2003 responses were coded (with 100 percent verification) by the Census Bureau using the 2000 Standard Occupational Classification System (SOC) and the 2002 North American Industry Classification System (NAICS). The 2002 NAICS is an updated and expanded version of the 4-digit Standard Industrial Classification (SIC) coding systems that was used to code CHIS 2001 responses. The newer NAICS uses a 6-digit coding system that improves the level of coding detail and expands the 9 divisions found in the SIC to more than 20 sectors in the NAICS. Westat reviewed the cases coded by the Census Bureau and posted them to the CHIS database.

Of all the random digit dialing (RDD) adult extended respondents (including both completed and partially completed interviews[5] contained in the data delivery file), about 60 percent answered the industry (AK5) and/or the occupation item (AK6). The Census Bureau staff coded more than 99 percent of the cases submitted for both the industry question and the occupation question. The few remaining cases were either blank or were reported as uncodeable.

[5] Adult extended interviews that are considered complete have disposition codes "CA" or "CP." CP includes all the partially completed adult interviews, i.e., interviews that were completed through Section I on the health insurance of the extended adult interview.

# 4. RACE AND ETHNICITY CODING

The procedures for coding the information collected in CHIS 2003 about respondent race and ethnicity remained virtually unchanged from the prior administration of CHIS in 2001. The survey items about race and ethnicity remained consistent with the data collected in the 2000 Census. This section describes how we handled situations when the respondent reported a race or ethnicity that was not classified into one of the pre-existing categories. These responses were recorded in the "other" category as a text string (the other specify). The procedures for coding these "other specify" responses into existing codes (up-coding) or leaving them in the other category are presented here.

Item AA5A from CHIS 2003 asked respondents for their race; "Please tell me which one or more of the following you would use to describe yourself. Would you describe yourself as Native Hawaiian, Other Pacific Islander, American Indian, Alaska Native, Asian, Black, African American, or White?" Although the question text and available response options were unchanged from CHIS 2001, the response options were reordered on the CATI screen to correct a problem discovered in CHIS 2001. In CHIS 2001, the race variable was coded as "Native Hawaiian" for a disproportionately large number of adult respondents. Through re-contact at that time, we learned that a number of these cases were miscoded by interviewers. Our hypothesis for the miscodes is that the response set for the race item was not ordered as in most studies. The first response option for most Westat CATI studies, is typically "white," not "Native Hawaiian" as was the case for CHIS 2001. The response options were read to respondents in the order as shown above in CHIS 2003, but presented in the more common order (white, black or African American, Asian, American Indian or Alaska Native, Other Pacific Islander, Native Hawaiian, or Other) for coding by interviewers.

The CHIS 2003 instrument allowed respondents to indicate that they identified with more than one race. The first question in the series of items related to race and ethnicity (question AA4 in the adult extended interview) asked if the respondent was Hispanic or Latino. If the response to this item was "yes," then a question was asked about the specific origin (Mexican, etc.) and this includes an "other" category with responses entered by interviewers as text in question AA5OS. The race question allowed the respondent to say they belonged to any or all of the coded races (Native Hawaiian, Other Pacific Islander, American Indian or Alaska Native, Asian, African American, or White) and the respondent could also say "other" race and that "other specify" race is recorded in text (AA5AOS). Another item follows if the respondent indicated they identified with more than one race or ethnicity. That item asked which race or ethnicity the respondent most identified with (AA5F). This item did not allow interviewers to collect an "other-specify," but responses to this item could be used in the coding decisions for other items.

## 4.1 Coding Procedures

The procedures for the race and ethnicity coding Westat performed were designed specifically to support the data needs for weighting the CHIS sample. If codes could not be assigned for race or ethnicity they were left as missing and were later imputed. The imputation procedures are described in *CHIS 2003 Methodology Series: Report 5 - Weighting and Variance Estimation*.

The procedures we used were consistent with the ones used to code the 2000 Census data and with those used in CHIS 2001. The methods used in the 2000 Census are documented in *Census 2000 Redistricting Data (Public Law 94-171) Summary File – Technical Documentation* (U.S. Census Bureau, 2001) available at http://www.census.gov/prod/www/abs/pl94-171.pdf. The specific sections of interest are in Appendix B, pages B-2 and B-3. When we refer to the Census procedures, we mean our interpretation of the information in this document.

An initial review of cases showed that the largest group of cases with "other race" categories were ones in which the respondent identified as being Hispanic or Latino and did not identify with any pre-coded race categories. The typical response to the "other race" was "Hispanic." Following the Census procedures, the person was left in the "other race" category and the "other specify" text remained as it was.

The specific procedures and guidelines we used are detailed below and are unchanged from those used in the past administration of the survey. Responses captured in the other specify text field were retained and included in the final data set delivery to UCLA to accommodate other research and analytic needs.

- If the "other specify" text clearly should have been included in an existing code (following the Census procedures), then it was up-coded and removed from the "other" category. For example, if the respondent was coded only as other race and the "other specify" was "Irish," then the code for "white" was upcoded to "yes," other race was revised to "no" and the other specify text eliminated.

- If the "other specify" text did not fit into an existing code (following the Census procedures), then it was left in the "other" category with the existing text in the "other specify." For example, if the "other specify" text for race was "Indian" and no other race category was identified, then no changes were made in the responses.

- If the respondent was coded as being Hispanic or Latino, we never revised this code based upon information in the other specify comments of the other variables. For example, if the person was coded as Hispanic and the origin item was only coded as "other" with the text "Jewish," then the Hispanic code was not altered.

- If the respondent was coded as not being Hispanic or Latino but the text in the "other specify" field for race indicated they were Hispanic or Latino, then the Hispanic or Latino coding was revised to "yes." In addition, the specific Hispanic origin code was made consistent with text in the "other specify" text from the race variable, if it was possible to do so. In the case where this was not possible, the "other" Hispanic origin category was coded and the text copied from the race variable to the "other specify" for Hispanic origin. (This procedure is an elaboration of the ones above to deal with the cross-variable coding.) For example, if the race "other specify" code was "Mexican," then the Hispanic or Latino category was revised to be "yes" and the Hispanic origin code was coded as "yes" for Mexican.

- If the "other race" text was similar to "none of above," we left the response as it was.

- If the "other race" text was similar to "human race," we coded this as a refusal. The race was then imputed along with other cases that were more direct refusals.

The Census procedures clearly state that persons who say they have European, Middle Eastern, or North African origin are to be classified as "White" race. This rule has many implications. For example, suppose a person says they are not Hispanic and only identify the "other race" as being Spain. We would upcode Hispanic origin to "yes" (to be consistent with the Census procedures for Hispanic origin) and then upcode "race" to "White" (since the person is of European origin).

## REFERENCES

U.S. Census Bureau. (2001). *Census 2000 Redistricting Data (Public Law 94-171) Summary File – Technical Documentation.*