



California  
Health  
Interview  
Survey

Making California's Voices Heard on Health

CHIS 2001 Methodology Series

Report 3

Data Processing Procedures

**CALIFORNIA HEALTH INTERVIEW SURVEY**

**CHIS 2001 METHODOLOGY SERIES**

**REPORT 3**

**DATA PROCESSING PROCEDURES**

**August 2002** *(rev 4/14/03)*

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



[www.chis.ucla.edu](http://www.chis.ucla.edu)

This report describes the data processing and editing procedures for CHIS 2001 that were performed by Westat, Inc. 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 2001 Methodology Series: Report 3 – Data Processing Procedures*. Los Angeles, CA: UCLA Center for Health Policy Research, 2002.

Copyright © 2002 by the Regents of the University of California.

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 2001 came from multiple sources: the California Department of Health Services, The California Endowment, the California Children and Families Commission, the National Cancer Institute, the Centers for Disease Control and Prevention, and the Indian Health Service.

## PREFACE

*Data Processing Procedures in CHIS 2001* is the third report in a series of methodological reports describing the 2001 California Health Interview Survey (CHIS 2001). The other reports are listed below. This report describes the data processing procedures that took place at the site of Westat, Inc. It does not include the additional processing procedures that later took place at UCLA. Please check the CHIS website ([www.chis.ucla.edu](http://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 the data collection and the preparation of five methodological reports from the 2001 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 2001 are as follows:

- Report 1: Sample Design for CHIS 2001
- Report 2: Data Collection Methods in CHIS 2001
- Report 3: Data Processing Procedures in CHIS 2001
- Report 4: Response Rates in CHIS 2001
- Report 5: Weighting and Variance Estimation for CHIS 2001

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 2001. One chapter details the data editing procedures and addresses the steps taken for ensuring data quality. Delivery of the final data sets are 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. The final chapter summarizes the follow-up activities related to cases in which the adult respondent was identified as a Native Hawaiian.

## TABLE OF CONTENTS

<u>Chapter</u>		<u>Page</u>
1	OVERVIEW .....	1-1
	1.1 Overview .....	1-1
	1.2 Sample Design Objectives.....	1-2
	1.3 Data Collection .....	1-3
	1.4 Response Rate.....	1-5
	1.5 Weighting the Random Digit Dial Sample.....	1-6
	1.6 Imputation Methods .....	1-8
	1.7 Methodology Report Series.....	1-9
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 .....	3-10
4	RACE AND ETHNICITY CODING .....	4-1
	4.1 Coding Procedures.....	4-1
5	FOLLOW-UP CALLS.....	5-1

### List of Tables

<u>Table</u>		
1-1	California county and county group strata used in the sample design.....	1-2
1-2	Number of completed interviews by type of sample, instrument.....	1-4
1-3	Survey topic areas by instrument .....	1-5

## TABLE OF CONTENTS

### List of Tables (continued)

<u>Table</u>		<u>Page</u>
2-1	Response codes added to CHIS 2001 .....	2-5
3-1	ZIP Codes within sampling stratum .....	3-4
3-2	Final distribution of adult extended completed cases by self-reported and original sampling stratum (including San Francisco and Santa Barbara) .....	3-11
5-1	Summary of final cases disposition .....	5-2

## **1. CHIS 2001 DESIGN AND METHODOLOGY SUMMARY**

### **1.1 Overview**

The 2001 California Health Interview Survey (CHIS 2001) is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. The focus of the survey is on a variety of public health topics, including access to health care and health insurance coverage. CHIS 2001 is the largest state health survey ever undertaken in the United States. It is a random digit dialing (RDD) telephone survey of California households designed to produce reliable estimates for the whole state, for large- and medium-sized population counties in the state, and for groups of the smallest population counties. Three California cities that have their own health departments were also sampled as part of CHIS 2001.

The survey design supports study of California's major race and ethnic groups, and a number of smaller ethnic groups within the state. Adults, parents of children below age 12, and adolescents (ages 12-17) residing in California households are the eligible respondents to the survey. CHIS 2001 collected data between November 2000 and October 2001. The plans are to conduct independent cross-sectional surveys of the California population on a biannual basis to monitor important health-related indicators and potentially track changes over time. CHIS 2001 is the first of these planned surveys.

CHIS 2001 collected information on if, where, and how people get health care in California. The goal is to provide health planners, policymakers, state, county, and city health agencies, and community organizations with information on the health and health care needs facing California's diverse population. For example, the number and characteristics of adults, children, and adolescents without access to care and lacking health insurance can be estimated from the data collected in CHIS 2001. Other key estimates on the prevalence of cancer screening, diabetes, asthma, and other health conditions can also be produced. The survey includes major content areas, such as health status and conditions, health-related behaviors, access to health care services, and health insurance coverage.

## 1.2 Sample Design Objectives

The CHIS 2001 sample is designed to meet two objectives: (1) provide local-level estimates for counties and groupings of counties with populations of 100,000 or more; and (2) provide statewide estimates for California’s overall population and its larger race/ethnic groups, as well as for several smaller ethnic groups. To address these objectives, the sample was allocated by county and aggregates of smaller counties, with supplemental samples of selected populations and cities. Table 1-1 shows the 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.

Table 1-1. California county and county group strata used in the 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, 2001 California Health Interview Survey.

Samples were also drawn from each of the three California cities that have their own local health department. In addition, supplemental samples were developed for three counties that contracted for additional sample to enhance their overall estimates. These city and supplemental county samples were in the following locations:

- The cities of Berkeley, Long Beach, and Pasadena; and
- The counties of San Francisco, Santa Barbara, and Solano.

The three city samples and the Solano county supplemental sample were implemented with and incorporated in the original statewide RDD sample. The separate San Francisco and Santa Barbara supplemental samples were subsequently added to the statewide RDD sample prior to constructing the sample weights and are part of the final CHIS 2001 RDD sample file.

To accomplish the second objective, larger sample sizes were allocated to the more urban counties where a significant portion of the state's African American and Asian ethnic populations reside. Additionally, supplemental samples were used to improve the sample size and precision of the estimates for specific ethnic groups. The supplemental ethnic group samples in CHIS 2001 were as follows:

- South Asian, Cambodian, Japanese, Korean, and Vietnamese;
- American Indian/Alaska Natives in urban and rural areas; and
- Latinos residing in Shasta County (a sample requested by the local health department).

### **1.3 Data Collection**

To capture the rich diversity of the California population, interviews were conducted in six languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, Korean, and Khmer (Cambodian). These languages were chosen based on research that identified these as the languages that would cover the largest number of Californians in the CHIS sample design 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 2001 data collection for the CHIS project. Westat staff interviewed one randomly selected adult in each sampled household. In those households with children (under age 12) or adolescents (ages 12-17), 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 who knew the most about the health and care of the sampled child was interviewed. The sampled adolescents responded for themselves, but only after a parent or guardian gave permission for the interview. Since adolescents were not reliable sources concerning their own health insurance coverage, the parents of sampled adolescents were interviewed about this topic separately.

One criterion for the adolescent and child to be selected for the survey is that they had to be “associated” with the selected adult. This meant that in most cases the interviewed adult had to be either the parent or guardian. The CHIS 2001 sample weights adjust for this selection criterion so as not to bias estimates based on the adolescent and child surveys. Table 1-2 shows the number of completed adult, child, adolescent, and adolescents’ health insurance interviews in CHIS 2001, by the type of sample (RDD or supplemental sample).

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

Type of sample	Adult	Child	Adolescent	Adolescent insurance
Total RDD + supplemental cases	57,848	13,276	6,058	8,302
RDD (includes 3 cities + Solano county supplemental cases)	54,122	12,392	5,733	7,809
Santa Barbara supplemental cases	206	49	22	31
San Francisco supplemental cases	1,100	151	46	79
<i>Total CHIS 2001 RDD file</i>	<i>55,428</i>	<i>12,592</i>	<i>5,801</i>	<i>7,919</i>
Other supplemental samples:				
South Asian	443	158	39	65
Cambodian	126	44	37	44
Japanese	330	51	18	33
Korean	326	95	30	44
Vietnamese	540	124	34	60
American Indian/Alaska Native	351	106	51	71
Shasta Latinos	304	106	48	66

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

The interviews done in English were administered using Westat’s computer-assisted telephone interviewing (CATI) system. Spanish and Vietnamese language interviews were also conducted entirely in CATI, while interviews conducted in Cantonese, Mandarin, Korean, and Khmer used English CATI screens and paper translations in tandem. The average adult interview took around 32 minutes to complete. The average child and adolescent interviews took 14 minutes and 19 minutes, respectively. Interviews in the non-English languages generally averaged longer to complete. Approximately 12 percent of the adult interviews were completed in a language other than English, as were 21 percent of all child (parent proxy) interviews and 9 percent of all adolescent interviews.

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

Table 1-3. Survey topic areas by instrument

Adult interview	Child interview	Adolescent interview
Age, sex, race, ethnicity	Age, sex, race, ethnicity	Age, sex, race, ethnicity
Physical activity		Physical activity
	Bike helmet use	Bike helmet, seatbelt use
	Recent serious injury	Recent serious injury
Health status	Health status	Health status
Women's health	Child care	
Chronic health conditions	Asthma, ADD	Asthma, diabetes
Cancer history, screening		
Skin cancer prevention	Skin cancer prevention	Skin cancer prevention
Health care use and access	Health care use and access	Health care use and access
Alcohol, tobacco use		Alcohol, tobacco, drug use
Mental health		Mental health
Health insurance	Health insurance	Health insurance
Diet (fruit-vegetable intake)	General diet	General diet
Dental health	Dental health	Dental health
Employment		Employment
Gun access, training		Gun access, violence
Income		
	Family interaction	Parental involvement
	Video games, computer use	Video games, computer use
Sexual orientation		Sexual behavior, orientation
		Future plans

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

## 1.4 Response Rate

The overall response rate for CHIS 2001 is a composite of the screener completion rate (i.e., success in introducing the survey to a household in order to select a respondent), and the extended interview completion rate (i.e., success in getting the selected respondent to complete the full interview). For the adult survey, the screener completion rate was 59.2 percent and the extended interview completion rate was 63.7 percent. This gives an overall response rate of 37.7 percent. To maximize the survey's response rate, an advance letter (in five languages) was mailed to all sampled telephone numbers for which an address could be obtained from reverse directory services. Approximately 66 percent of the sample was mailed an advance letter. Response rates varied by sampling stratum and were slightly higher in households that received an advance letter.

To assist in achieving sample size goals, respondents that completed 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.” This resulted in 397 “partial completes” being included in the final adult survey data. Employment and income information as well as potential public program eligibility and food insecurity information would be missing from these cases.

Proxy interviews were allowed for frail and ill persons over the age of 65. The reason is that health estimates made for elderly persons could be biased if this is not allowed. Eligible selected persons were recontacted and offered a proxy option and 316 had a proxy interview completed by either a spouse/partner or adult child. Only a subset of questions identified as appropriate for a proxy respondent were administered.

## **1.5 Weighting the Random Digit Dial Sample**

To produce correct 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. Sample weighting was carried out in CHIS 2001 to accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons (Note: households with listed addresses and thus eligible for an advance letter were assigned a probability of selection of 1.25 over unlisted households);
- 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 for the RDD samples (each stratum is an independent sample), a household weight was created for all households that completed the screener interview. This household weight is the “base weight” computed as the inverse of the probability of selection of the sample telephone number adjusted for each of the following:

- Subsampling for listed address/advance letter status;
- Unknown residential status;
- Screener interview nonresponse;
- Multiple telephone numbers; and
- Household poststratification.

A “poststratified household weight” was then used to compute a person-level weight. This person-level weight incorporates the within-household probability of selection of the sampled person and adjusts for nonresponse, plus an adjustment resulting from raking the data to person-level control totals. Each of these adjustments corresponds to a multiplicative weighting factor.

Raking can be thought of as a multidimensional poststratification procedure because the weights are basically poststratified to one set of control totals (a dimension), then these adjusted weights are poststratified to another dimension. After all dimensions were adjusted, the process was iterated until the control totals for all the dimensions were simultaneously satisfied (within a specified tolerance).

There are 11 dimensions used in CHIS 2001. The first 10 dimensions are created by combining demographic variables (age, sex, race, and ethnicity) and different geographic areas (city, county, group of counties, and state). The 11th dimension is created to adjust the weights for households without a telephone number.

The control totals used in the raking were derived from the *Census 2000 Summary File 1* (SF1). Population items in SF1 include sex, age, race, ethnicity (Latino/non-Latino), household relationships, and group quarters. The race classification in SF1 include six groups: White, African American, American Indian/Alaska Native, Asian, Native Hawaiian/Pacific Islander, and a category of Other Race. Since a person could report multiple races, the SF1 provided counts for each of 63 possible race combinations a person could report.

One of the limitations of using the SF1 for the control totals is the inability to produce counts that exclude the fraction of the population living in “group quarters” (e.g., nursing homes, prisons) for some dimensions used in CHIS 2001. The group quarter population represented 2.4 percent of the total population in California. As a result, the number of persons living in group quarters was estimated for some of the raking dimensions, and the SF1 totals were reduced by these estimated amounts prior to raking.

## **1.6 Imputation Methods**

Three different imputation procedures were used in CHIS 2001 to fill in missing responses that were essential for weighting the data or for such basic descriptive purposes as income categories. The first imputation technique is deterministic or non-stochastic in nature. Deterministic imputation was used to fill in the missing items for self-reported county of residence (item AH42). These imputations required no randomization because other geographic data are available that can be used to determine the respondent’s county of residence with a relatively high level of probability of being correct although not with 100 percent certainty in all cases.

The second imputation technique is a completely random selection from the observed distribution. This method is used only when a very small percentage of the items are missing. For example, when imputing the missing values for self-reported age, 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 third technique is hotdeck imputation. Hotdeck imputation was used to impute race, ethnicity, and household income in CHIS 2001. The hotdeck approach is probably the most commonly used method for assigning values for missing responses in large-scale household surveys.

With a hotdeck, 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. To carry out hotdeck imputation for CHIS 2001, 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, with the same household structure. The recipient is then randomly imputed the same household income, ethnicity/race (depending on the items that need to be imputed) from one of the donors in the pool. Once a donor is used, it is removed from the pool of donors.

Imputation flags are used in the data file 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 2001:

- Report 1 - Sample Design
- Report 2 - Data Collection Methods
- Report 3 - Data Processing Procedures
- Report 4 - Response Rates
- 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](http://www.CHIS.ucla.edu) or contact CHIS at [CHIS@ucla.edu](mailto:CHIS@ucla.edu).

## 2. DATA EDITING PROCEDURES

CHIS 2001 survey data 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. (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 is programmed to determine which questions are asked based on 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, careful quality control procedures were implemented. CHIS 2001 survey data were carefully examined and edited before delivering final data files. These procedures involved limiting the number of staff who made updates, using the CATI specifications for resolving 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 2001 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.

During data collection, an interviewer who experienced a problem while working a case 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 telephone facilities were sent to a single staff member at the Westat TRC who was responsible for distributing them to the appropriate Westat 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 code the appropriate result for the case without completing a paper problem sheet. These cases were assigned a problem result code in the CATI system. The result code had three subcodes for special queues to which these problem cases could be assigned for review. These subcodes were used to indicate the person responsible for investigating the case further—TRC staff, project staff, or programmers. Such cases were reviewed electronically by a TRC supervisor and either re-fielded to the interviewers or distributed to the TRC, programming, or project staff as appropriate.

For some problems, database updates were unnecessary and the cases could simply be released for general interviewing using an appropriate message. If, for example, a case was stopped during the middle of Section D of the adult extended interview, the interviewer would type a detailed comment telling 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 AD1.” Most restart cases were made available to the general interviewing staff. For highly unusual or extremely complex cases, the case could be assigned to a specific interviewer with experience in these types of cases.

When cases were changed, they were often re-released for general interviewing. Some examples of these types 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 child needed changing. 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 with a project sampling statistician and making the appropriate changes to the data or refielding 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 questionnaire. For CHIS 2001, sometimes these phrases were merely an elaboration of a previously recorded response or an expression of opinion. Other times, they 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. For example, question AD12 from the adult extended interview asked female respondents; “Have you had a hysterectomy?” The CATI response options were “Yes” and “No.” Respondents’ answers included, “I’ve had a partial hysterectomy,” and “I’ve had one ovary removed, does this count?” Once the response was clarified it was applied to other similar cases.

Another question that elicited several responses outside the standard response set was AB34 from the adult extended interview, “Has a doctor every told you that you have any kind of heart disease?” Interviewer comments for this item included: “Respondent reported having a leaky valve,” “heart arrhythmia,” and “respondent has an irregular heart beat.” Westat project staff consulted with UCLA

project staff to accurately code such responses and the interviewing staff was informed how to code these responses.

Data preparation and project staff met weekly 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**

For most items in CHIS 2001 survey, closed-ended response options were provided and no coding was needed. A small number of other-specify type questions did require coding. 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 AA5A in the adult extended interview asked respondents “Also, 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?” An “other” category was available for responses that fell outside the list of races read as a part of the question. Additional questions with an other category from CHIS 2001 adult extended interview included ethnic ancestry questions (AA5, AA5E, AA5E1), tribal names (AA5B, AA5D), place visited for health care (AH3, AH7), country in which born (AH33, AH34, AH35), languages spoken at home (AH36), and health insurance coverage items (AI15, AI17A, AI24).

Westat data preparation staff reviewed the “other” responses and up-coded them to the existing categories whenever possible. Response codes were added for some survey items to accommodate the answers recorded in the other-specify category. The updated response codes for these items are given in Table 2-1.

Table 2-1. Response codes added to CHIS 2001

Questionnaire version	Item	New code	Response description
Adult	AD1	96	Never started menstrual cycle
Adult	AE22	4	Does a combination of two or more of the items listed equally
		5	Lays down most of the time
Adult	AF9	4	Never go out in the sun for more than 1 hour/never go out in the sun.
	AF10	4	Never go out in the sun for more than 1 hour/never go out in the sun.
	AF11	4	Never go out in the sun for more than 1 hour/never go out in the sun.
	AF12	4	Never go out in the sun for more than 1 hour/never go out in the sun.
Teen	TE16	4	Never go out in the sun for more than 1 hour/never go out in the sun.
Child	CB9	4	Never go out in the sun for more than 1 hour/never go out in the sun.
Adult	AH47	30	Had no formal education
Child	CH22	30	Had no formal education
Child	CH26	30	Had no formal education
Child	CH29	30	Had no formal education
Adult	AH47	30	Had no formal education
Adult	AI15	10	Covered under same plan spouse/same company
		11	Doesn't like/want company insurance
Adult	AI36	13	Had insurance all 12 months, just now lost
		14	Denied coverage, not specified/doesn't qualify not specified
		15	Do have coverage but don't know type
		16	Switched insurance companies, delay between
		17	Didn't like insurance offered/didn't want it
Adult	AI24	13	Had insurance all 12 months, just now lost
		14	Denied coverage, not specified/doesn't qualify not specified
		15	Do have coverage but don't know type
		16	Switched insurance companies, delay between
		17	Didn't like insurance offered/didn't want it

Table 2-1. Response codes added to CHIS 2001 (continued)

Questionnaire version	Item	New code	Response description
Teen insurance	IA18	13	Had insurance all 12 months, just now lost
		14	Denied coverage, not specified/doesn't qualify not specified
		15	Do have coverage but don't know type
		16	Switched insurance companies, delay between
		17	Didn't like insurance offered/didn't want it
Teen insurance	IA29	13	Had insurance all 12 months, just now lost
		14	Denied coverage, not specified/doesn't qualify not specified
		15	Do have coverage but don't know type
		16	Switched insurance companies, delay between
		17	Didn't like insurance offered/didn't want it
Child	CF29	13	Had insurance all 12 months, just now lost
		14	Denied coverage, not specified/doesn't qualify not specified
		15	Do have coverage but don't know type
		16	Switched insurance companies, delay between
		17	Didn't like insurance offered/didn't want it
Child	CF18	13	Had insurance all 12 months, just now lost
		14	Denied coverage, not specified/doesn't qualify not specified
		15	Do have coverage but don't know type
		16	Switched insurance companies, delay between
		17	Didn't like insurance offered/didn't want it
Teen	TH6A	4	Never goes out at night
Teen insurance	IA32	3	Mother/father deceased
	IA35	3	Mother/father deceased
Child	CH13	3	Mother/father deceased
	CH16	3	Mother/father deceased

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

With the exception of the industry and occupation items in the adult extended interview, CHIS 2001 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.

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. This enabled errors to be identified and corrected with the respondent without requiring a call back.

Soft-range edits were activated during the interview when the respondent gave an improbable response (a value outside the specified range). The CATI system responded by placing a message on the screen and requiring that the interviewer 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. 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 edit specifications, interviewers recorded the respondent's answer in the comment field and data preparation staff 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 from problem sheets so that the final survey data reflected the most accurate information possible.

Survey updates were verified using a variety of techniques. To ensure accuracy, intended updates were first recorded on a hard-copy printout or on an associated problem sheet. This printout was

checked for accuracy and for logical effects on other questions or skip patterns in the questionnaire. The updates were then entered into the computer and checked again. For cases that were difficult to resolve, other information was reviewed, such as interviewer comments, messages, and problem descriptions, in order to ensure proper handling of the update.

An electronic transaction journal recorded each update and tracked information such as the 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. We estimate that approximately 15,000 to 20,000 database values were updated and verified.

Cases with similar problems were reviewed together and then updated at one time in manageable batches. This process ensured consistency in the handling of a discrete data problem type. Following the series of updates, a program was run to check for the full set of errors that had been identified at that point in time to ensure that data editing had not created any new errors. Frequency distributions and cross-tabulations 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, the problem case was identified, 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**

In advance of delivering the final survey data sets to UCLA, the CATI database was converted to SAS data sets. Essentially, the CATI data are stored in a hierarchical database to improve data efficiency and enhance performance while interviewing. The SAS “flat” files were prepared for analytic purposes. This conversion was accomplished using Westat’s CATISAS macro program that extracted information stored in the CATI data dictionary (e.g., variable names, variable labels, allowable

values, and formats) and then converted each of the CATI database segments into a SAS data set. Using the data dictionary to define the SAS data sets variables was 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.

After each major stage in the conversion process 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; geocodes, and Standard Industrial Classification/Standard Occupational Code;
- Weights (final sample weight and replicate weights); and
- Imputation flags.

### **3. GEOGRAPHIC AND INDUSTRY AND OCCUPATION CODING**

For CHIS 2001, Westat delivered coded survey data for items related to where the adult respondent lived, and also for items related to industry and occupation. This report section discusses the procedures Westat used for the geographic coding and also provides a summary of the industry and occupation coding performed by the U.S. Census Bureau.

#### **3.1 Geographic Coding**

The adult extended interview asked respondents who reported living in San Diego or Los Angeles county to provide the name of the street where they lived (question AM8) and the name of a cross street near their residence (question AM9). All cases from the adult extended interviews that had information for either location question (n=12,654) were selected for coding.

Geocoding CHIS 2001 data was both an iterative and interactive process that utilized specialized software packages and manual review. The process started with matching the street location and ZIP Code survey data to relevant information in “Streetmap 2000” using Environmental Systems Research Institute Inc.’s (ESRI’s) “Arcview” software. We conducted this first step in “batch” mode to reduce the burden associated with interactive searching. Using a combination of state (California), ZIP Code (question AM7) and the street names collected in the adult extended interview as input values, the software attempts to resolve these data values and return a specific location, or ZIP Code area.

After batch processing, we performed interactive searches and look-ups for all non-matching records. This work was made more demanding because some street locations did not intersect, some streets were nonexistent, street names were misspelled, as noted above, the ZIP Code information (question AM7) was not present for all cases. In addition, we noted some differences between the input and output street names because there are streets that have more than one name or alias. We corrected obvious spelling mistakes found in the street names and corrected ZIP Code whenever possible to maximize the number of matches.

In interactive mode, the software returned a ZIP Code associated with where the point of interest (i.e., the intersection of the two street locations) is located. Next, a comparison was made between

the input ZIP Code (ZIP Code from question AM7) and the ZIP Code returned by the software. When the two ZIP Codes do not match, this means the location falls on a ZIP Code boundary and the software randomly selected one of the two ZIP Codes. There were 478 cases in which the two ZIP Codes did not match. For cases that did not result in a match after this initial interactive processing, we attempted to find additional matches by using three additional sources including Mapquest, GDT, and Mapblast.

Next, for all the cases with a ZIP Code match, the identified ZIP Code was overlaid against ESRI's database of the 1990 California ZIP Code polygons and against a national ZIP Code point database. At the time of processing, there were 1,646 ZIP Code polygons in California (originating from the 1990 Census data.) and 41,738 national ZIP Codes. Results from this process yielded an output database containing the original ZIP Code from question AM7, the geocoded ZIP Code, and a polygon ZIP Code. After several interactive passes through the data the final match rate was 85.5 percent (n=10,814). The year 2000 Census ZIP Code polygon information has subsequently become available and if used, may result in a higher success rate.

As described in Report 1: Sample Design, the 41 sampling strata for CHIS 2001 were principally constructed at the county level. This design also included three "city" strata corresponding to Pasadena and Long Beach in Los Angeles county and Berkeley in Alameda county. Because the sampling strata are assigned based on the telephone exchange and some of the exchanges serve more than one county or city, the actual stratum where the sampled person resides may differ from the sampling stratum. 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.

Two questions from the adult extended interview were used to make the stratum assignment and to construct a variable representing the respondent's self-reported stratum. 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 is derived from the self-reported (imputed when missing) county values (AH42) for all strata except those strata corresponding to the cities of Pasadena, Long Beach and Berkeley. For these city strata, the self-reported stratum assignment is based on self-reported ZIP Code

(AM7). Using information from the Census about geographical areas and their associated ZIP Codes, the self-reported ZIP Code was then used in making an assignment to the appropriate city stratum. Table 3-1 shows a listing of ZIP Codes within each stratum.

Not all respondents answered the item about the county in which they lived or provided their ZIP Code, however. For individuals who did not report their ZIP Code, this item was imputed prior to making the assignment to the self-reported stratum. This imputation process took advantage of auxiliary information available from the geo-coding effort and also from the telephone exchange data.

**Auxiliary Information.** The missing ZIP Code information noted above was augmented by using the resultant ZIP Code from the geo-coding procedures or by using the ZIP Code associated with the sampled telephone number. Recall that adult respondents who reported living in either Los Angeles or San Diego County were also asked for the name of the street where they lived (AM8) and the name of a near-by cross-street (AM9). This self-reported street location information was geo-coded and as a result, one or more ZIP Codes were identified for most of these cases (85 percent success rate).

Additional ZIP Code information was available from the sampling vendor and from companies used to match sampled telephone numbers to street addresses. (Street addresses were identified so that a prenotification letter announcing CHIS could be mailed. See Report 3: Data collection, for a detailed discussion.) The sampling vendor provided as many as 6 ZIP Codes for the sampled telephone exchanges.

**Stratum Assignment.** Given these several data sources, a decision hierarchy was established for making assignments to the self-reported sampling stratum. When the self-reported ZIP Code and county information from the adult extended interview was internally consistent (that is, the self-reported ZIP Code was located within the self-reported county), the case was assigned to that stratum, even if this stratum differed from the original sampling stratum. If either the self-reported ZIP Code or county was not answered, the non-missing response was used to identify the appropriate stratum.

Table 3-1. ZIP Codes within sampling stratum

	Stratum	Zip Code
1.1	Long Beach	90745, 90746, 90747, 90749, 90801, 90802, 90803, 90804, 90805, 90806, 90807, 90808, 90809, 90810, 90813, 90814, 90815, 90822, 90831, 90832, 90833, 90834, 90835, 90840, 90842, 90844, 90845, 90846, 90847, 90848, 90853, 90888, 90899
1.2	Pasadena	91050, 91051, 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
1.3	Remainder of 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, 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, 90748, 91001, 91003, 91006, 91007, 91009, 91010, 91011, 91012, 91016, 91017, 91020, 91021, 91023, 91024, 91025, 91030, 91031, 91040, 91041, 91042, 91043, 91046, 91066, 91077, 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, 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, 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, 93510, 93532, 93534, 93535, 93536, 93539, 93543, 93544, 93550, 93551, 93552, 93553, 93563, 93584, 93586, 93590, 93591, 93599

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, 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, 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, 90680, 90720, 90721, 90740, 90742, 90743, 92602, 92603, 92604, 92605, 92606, 92607, 92609, 92610, 92612, 92614, 92615, 92616, 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, 95002, 95008, 95009, 95011, 95013, 95014, 95015, 95020, 95021, 95026, 95030, 95031, 95032, 95035, 95036, 95037, 95038, 95042, 95044, 95046, 95050, 95051, 95052, 95053, 95054, 95055, 95056, 95070, 95071, 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, 91784, 91785, 91786, 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, 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, 92397, 92398, 92399, 92401, 92402, 92403, 92404, 92405, 92406, 92407, 92408, 92410, 92411, 92412, 92413, 92414, 92415, 92418, 92420, 92423, 92424, 92427, 93558, 93562, 93592
6	Riverside	91752, 92201, 92202, 92203, 92210, 92211, 92220, 92223, 92225, 92226, 92230, 92234, 92235, 92236, 92239, 92240, 92241, 92253, 92254, 92255, 92258, 92260, 92261, 92262, 92263, 92264, 92270, 92274, 92276, 92282, 92292, 92320, 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	Berkeley	94701, 94702, 94703, 94704, 94705, 94706, 94707, 94708, 94709, 94710, 94712, 94720
7.2	Remainder of Alameda	94501, 94502, 94536, 94537, 94538, 94539, 94540, 94541, 94542, 94543, 94544, 94545, 94546, 94550, 94551, 94552, 94555, 94557, 94560, 94566, 94568, 94577, 94578, 94579, 94580, 94586, 94587, 94588, 94601, 94602, 94603, 94604, 94605, 94606, 94607, 94608, 94609, 94610, 94611, 94612, 94613, 94614, 94615, 94617, 94618, 94619, 94620, 94621, 94622, 94623, 94624, 94625, 94626, 94627, 94643, 94649, 94659, 94660, 94661, 94662, 94666
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, 95608, 95609, 95610, 95611, 95615, 95621, 95624, 95626, 95628, 95630, 95632, 95638, 95639, 95641, 95652, 95655, 95660, 95662, 95670, 95671, 95673, 95680, 95683, 95690, 95693, 95741, 95742, 95743, 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, 95873, 95887, 95894, 95899
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, 94553, 94556, 94561, 94563, 94564, 94565, 94569, 94570, 94572, 94575, 94583, 94595, 94596, 94597, 94598, 94801, 94802, 94803, 94804, 94805, 94806, 94807, 94808, 94820, 94850

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

	Stratum	Zip Code
10	Fresno	93210, 93234, 93242, 93602, 93605, 93606, 93607, 93608, 93609, 93611, 93612, 93613, 93616, 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, 93762, 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, 94159, 94160, 94161, 94162, 94163, 94164, 94165, 94166, 94167, 94168, 94169, 94170, 94171, 94172, 94175, 94177, 94188
12	Ventura	91319, 91320, 91358, 91359, 91360, 91361, 91362, 91377, 93001, 93002, 93003, 93004, 93005, 93006, 93007, 93009, 93010, 93011, 93012, 93015, 93016, 93020, 93021, 93022, 93023, 93024, 93030, 93031, 93032, 93033, 93034, 93035, 93040, 93041, 93042, 93043, 93044, 93060, 93061, 93062, 93063, 93064, 93065, 93066, 93093, 93094, 93099
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, 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, 93380, 93381, 93382, 93383, 93384, 93385, 93386, 93387, 93388, 93389, 93390, 93501, 93502, 93504, 93505, 93516, 93518, 93519, 93523, 93524, 93527, 93528, 93531, 93554, 93555, 93556, 93560, 93561, 93581, 93596
15	San Joaquin	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, 95290, 95296, 95297, 95298, 95304, 95320, 95330, 95336, 95337, 95366, 95376, 95377, 95378, 95385, 95391, 95686
16	Sonoma	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	95307, 95313, 95316, 95319, 95323, 95326, 95328, 95329, 95350, 95351, 95352, 95353, 95354, 95355, 95356, 95357, 95358, 95360, 95361, 95363, 95367, 95368, 95380, 95381, 95382, 95386, 95387, 95390, 95397

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

	Stratum	Zip Code
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, 93254, 93427, 93429, 93434, 93436, 93437, 93438, 93440, 93441, 93454, 93455, 93456, 93457, 93458, 93460, 93463, 93464
19	Solano	94510, 94512, 94533, 94535, 94571, 94585, 94589, 94590, 94591, 94592, 95620, 95625, 95687, 95688, 95696
20	Tulare	93201, 93207, 93208, 93218, 93219, 93221, 93223, 93227, 93235, 93237, 93244, 93247, 93256, 93257, 93258, 93260, 93261, 93262, 93265, 93267, 93270, 93271, 93272, 93274, 93275, 93277, 93278, 93279, 93282, 93286, 93290, 93291, 93292, 93603, 93615, 93618, 93633, 93647, 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	93401, 93402, 93403, 93405, 93406, 93407, 93408, 93409, 93410, 93412, 93420, 93421, 93422, 93423, 93424, 93428, 93430, 93432, 93433, 93435, 93442, 93443, 93444, 93445, 93446, 93447, 93448, 93449, 93451, 93452, 93453, 93461, 93465, 93483
24	Placer	95602, 95603, 95604, 95631, 95648, 95650, 95658, 95661, 95663, 95677, 95678, 95681, 95701, 95703, 95713, 95714, 95715, 95717, 95722, 95736, 95746, 95747, 95765, 96140, 96141, 96143, 96145, 96146, 96148
25	Merced	93620, 93635, 93661, 93665, 95301, 95303, 95312, 95315, 95317, 95322, 95324, 95333, 95334, 95340, 95341, 95342, 95343, 95344, 95348, 95365, 95369, 95374, 95388
26	Butte	95914, 95916, 95917, 95926, 95927, 95928, 95929, 95930, 95938, 95940, 95941, 95942, 95948, 95954, 95958, 95965, 95966, 95967, 95968, 95969, 95973, 95974, 95976, 95978
27	Shasta	96001, 96002, 96003, 96007, 96008, 96011, 96013, 96016, 96017, 96019, 96022, 96028, 96033, 96040, 96047, 96049, 96051, 96056, 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, 95937
29	El Dorado	95613, 95614, 95619, 95623, 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	92222, 92227, 92231, 92232, 92233, 92243, 92244, 92249, 92250, 92251, 92257, 92259, 92266, 92273, 92275, 92281, 92283
31	Napa	94503, 94508, 94515, 94558, 94559, 94562, 94567, 94573, 94574, 94576, 94581, 94599
32	Kings	93202, 93204, 93212, 93230, 93232, 93239, 93245, 93246, 93266
33	Madera	93601, 93604, 93610, 93614, 93637, 93638, 93639, 93643, 93644, 93645, 93653, 93669

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

	Stratum	Zip Code
34	Monterey, San Benito	93426, 93450, 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, 95023, 95024, 95039, 95043, 95045, 95075
35	Del Norte, Humboldt	95501, 95502, 95503, 95511, 95514, 95518, 95519, 95521, 95524, 95525, 95526, 95528, 95531, 95532, 95534, 95536, 95537, 95538, 95540, 95542, 95543, 95545, 95546, 95547, 95548, 95549, 95550, 95551, 95553, 95554, 95555, 95556, 95558, 95559, 95560, 95562, 95564, 95565, 95567, 95569, 95570, 95571, 95573, 95589
36	Lassen, Modoc, Siskiyou, Trinity	95527, 95552, 95563, 95568, 95595, 96006, 96009, 96010, 96014, 96015, 96023, 96024, 96025, 96027, 96031, 96032, 96034, 96037, 96038, 96039, 96041, 96044, 96046, 96048, 96050, 96052, 96054, 96057, 96058, 96064, 96067, 96068, 96085, 96086, 96091, 96093, 96094, 96097, 96101, 96104, 96108, 96109, 96110, 96112, 96113, 96114, 96115, 96116, 96117, 96119, 96121, 96123, 96127, 96128, 96130, 96132, 96134, 96136, 96137
37	Lake, Mendocino	95410, 95415, 95417, 95418, 95420, 95422, 95423, 95424, 95426, 95427, 95428, 95429, 95432, 95435, 95437, 95443, 95445, 95449, 95451, 95453, 95454, 95456, 95457, 95458, 95459, 95460, 95461, 95463, 95464, 95466, 95468, 95469, 95470, 95481, 95482, 95485, 95488, 95490, 95493, 95494, 95585, 95587
38	Colusa, Glen, Tehama	95912, 95913, 95920, 95931, 95932, 95939, 95943, 95950, 95951, 95955, 95963, 95970, 95979, 95987, 95988, 96021, 96029, 96035, 96055, 96059, 96061, 96063, 96074, 96075, 96078, 96080, 96090, 96092
39	Sutter, Yuba	95659, 95668, 95674, 95676, 95692, 95901, 95903, 95918, 95919, 95922, 95925, 95935, 95953, 95957, 95961, 95962, 95972, 95977, 95981, 95982, 95991, 95992, 95993
40	Plumas, Nevada, Sierra	95712, 95724, 95728, 95910, 95915, 95923, 95924, 95934, 95936, 95944, 95945, 95946, 95947, 95949, 95956, 95959, 95960, 95971, 95975, 95980, 95983, 95984, 95986, 96020, 96103, 96105, 96106, 96111, 96118, 96122, 96124, 96125, 96126, 96129, 96135, 96160, 96161, 96162
41	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	92328, 92384, 92389, 93512, 93513, 93514, 93515, 93517, 93522, 93526, 93529, 93530, 93541, 93542, 93545, 93546, 93549, 93623, 95221, 95222, 95223, 95224, 95225, 95226, 95228, 95229, 95232, 95233, 95245, 95246, 95247, 95248, 95249, 95250, 95251, 95252, 95254, 95255, 95257, 95305, 95306, 95309, 95310, 95311, 95314, 95318, 95321, 95325, 95327, 95335, 95338, 95345, 95346, 95347, 95364, 95370, 95372, 95373, 95375, 95379, 95383, 95389, 95601, 95629, 95640, 95642, 95644, 95646, 95654, 95665, 95666, 95669, 95675, 95685, 95689, 95699, 96107, 96120, 96133

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

In cases when the self-reported ZIP Code and the self-reported county did not point to the same stratum, the auxiliary information was used to determine the stratum assignment. The auxiliary information was rank-ordered such that the geo-coded data was considered the most reliable and accurate, followed by the information provided by the sampling vendor, and then the address information. When both the self-reported ZIP Code and self-reported county were missing, the same decision hierarchy was used.

This decision hierarchy was especially useful in resolving cases when the self-reported data seemed to fall on a ZIP Code boundary or when two sources appeared equally convincing. There were also a small number of cases in which the self-reported responses did not identify a single stratum and the auxiliary information was sufficiently ambiguous so that an assignment to the self-reported stratum was unclear. For these cases, each of the auxiliary data sources was reviewed in combination with geographical maps in effort to make the final stratum assignment.

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. 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 in which they live.

### **3.2 Industry and Occupation Coding**

CHIS 2001 adult extended interview included three questions about occupation and industry. These three questions were:

AK4 “On your main job, are you employed by: a private company; a Federal, state, or local 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?”

Table 3-2. Final distribution of adult extended completed cases by self-reported and original sampling stratum (including San Francisco and Santa Barbara)

Stratum name	Sampling stratum			Final self-reported stratum count
	count	Removed	Added	
1.1 - Long Beach	819	72	166	913
1.2 - Pasadena	814	178	35	671
1.3 - Remainder of Los Angeles	10,582	263	293	10,612
2 - San Diego	2,666	7	13	2,672
3 - Orange	2,495	57	16	2,454
4 - Santa Clara	1,514	29	25	1,510
5 - San Bernardino	1,547	30	37	1,554
6 - Riverside	1,386	17	22	1,391
7.1 - Berkeley	794	48	63	809
7.2 - Remainder of Alameda	1,191	90	64	1,165
8 - Sacramento	1,238	29	22	1,231
9 - Contra Costa	1,199	41	56	1,214
10 - Fresno	1,041	5	17	1,053
11 - San Francisco	1,969	29	14	1,954
12 - Ventura	971	3	47	1,015
13 - San Mateo	947	23	47	971
14 - Kern	1,096	6	3	1,093
15 - San Joaquin	1,052	7	13	1,058
16 - Sonoma	771	2	7	776
17 - Stanislaus	819	34	9	794
18 - Santa Barbara	1,004	7	4	1,001
19 - Solano	1,587	41	7	1,553
20 - Tulare	827	10	9	826
21 - Santa Cruz	793	10	8	791
22 - Marin	752	6	9	755
23 - San Luis Obispo	799	3	11	807
24 - Placer	784	34	14	764
25 - Merced	832	9	26	849
26 - Butte	825	2	12	835
27 - Shasta	826	4	5	827
28 - Yolo	834	7	17	844
29 - El Dorado	780	8	35	807
30 - Imperial	798	5	1	794
31 - Napa	806	4	31	833
32 - Kings	843	10	4	837
33 - Madera	824	6	2	820
34 - Monterey, San Benito	790	5	9	794
35 - Del Norte, Humbolt	861	13	7	855
36 - Lassen, Modoc, Siskiyou, Trinity	846	16	11	841
37 - Lake, Mendocino	813	8	3	808
38 - Colusa, Glen, Tehama	839	4	4	839

Table 3-2. Final distribution of adult extended completed cases by self-reported and original sampling stratum (including San Francisco and Santa Barbara) (continued)

Stratum name	Sampling stratum			Final self-reported stratum count
	count	Removed	Added	
39 - Sutter, Yuba	822	21	0	801
40 – Plumas, Nevada, Sierra	814	5	15	824
41 - Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	818	10	5	813
Total	55,428	1,218	1,216	55,428

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Respondents who reported working for the Federal, state, or local government in question AK4, were not asked the business or industry question (AK5). All respondents who indicated that they were working, however, were 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 to do the coding 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. Westat reviewed the coded cases and posted them to the CHIS database.

Of all the random digit dialing (RDD) adult extended respondents (including both completed and partially completed interviews<sup>1</sup> contained in the data delivery file), over 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 uncodeable.

<sup>1</sup> 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

In CHIS 2001, several data items collected information about the respondent's race and ethnicity. These items were structured to provide results consistent with the data collected in the 2000 Census and allow for respondents to indicate that they identified with multiple races. 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.

An item of significant interest in this process was the question about Hispanic or Latino origin (question AA4 in the adult extended interview). If the response to this item was "yes," then a question was asked about the origin (Mexican, etc.) and this includes an "other" category for which the "other specify" responses were recorded in text (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/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 have an "other specify" response that needed coding, 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 Report 5 (Weighting and Variance Estimation)

The procedures we used were consistent with the ones used to code the 2000 Census data. The methods used in the 2000 Census are documented in *Census 2000 Redistricting Data (Public Law 94-171) Summary File – Technical Documentation*.

The report is available at the Census Bureau URL: [www.census.gov/prod/www/abs/pl94-171.pdf](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. Note that we retained all of the original other specify text in a field delivered to UCLA so other analytic needs can be addressed.

- 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).

## 5. FOLLOW-UP CALLS

An examination of a preliminary data delivery showed that there were a disproportionately large number of adult respondents whose race was recorded as Native Hawaiian (code “1” in question AA5A). Although a careful review of the CATI instrumentation showed no error, staff were concerned about this result and a review these cases was undertaken. A soft range edit was added to question AA5A requiring interviewers to confirm all entries of “Native Hawaiian.”

After the review was conducted and discussed with UCLA, a determination was made to re-contact 137 households in which the adult respondent or other household members were coded as Native Hawaiian before the soft range edit was added. Nearly all of these 137 cases were finalized within the first 6 weeks of data collection. Due to the specialized and potentially sensitive nature in re-contacting these households, a group of highly experienced interviewers was selected for this work. These interviewers were also experienced in contacting households that had initially refused to participate in the study and were familiar with refusal avoidance techniques. Although interviewers did not need to re-administer the entire survey, we felt that this skill set was needed for re-asking respondents about their race and to avoid accepting refusals.

A separate 1-hour training session was conducted for this group of interviewers during the first week of June 2001. Interviewers were given instructions about the administrative aspects of their job as well as the contact procedures that they were expected to follow. Supporting information such as the case ID number, telephone number, month in which the initial interview was completed, and the name(s) of the adult household members who were coded as Native Hawaiian were provided for each of the 137 cases. This information was obtained directly from the CATI database.

For this small number of cases, printed “call-records” were used. Call-record forms were designed so interviewers could code the result of each telephone attempt made, and make notations that are pertinent to the case. The value of detailed notes cannot be understated because it was possible for more than one interviewer to work the same case at a later date.

Not unlike interviews conducted for the main study, a significant number of call attempts were made for each case in effort to properly identify the respondent’s race. On average, about six call attempts were made. The distribution of results from the follow-up interviews is presented in Table 5-1.

These results suggest that the initial coding for race in these cases was inaccurate. Of the 137 total cases, only 33 adult respondents verified their race as Native Hawaiian. A final determination about the respondent's race was not achieved for the 23 cases classified as refusal, no such person, nonworking telephone number, or no answer. One hypothesis for the relatively large number of miscodes is that the response set for the race item for CHIS 2001 was not ordered as in most studies. Consequently, experienced interviewers may have erred in coding responses because they were unfamiliar with the response order. In fact, for most CATI studies conducted by Westat, the first response option is typically "white," not "Native Hawaiian" as was the case for CHIS. We should also note, however, that the response categories were purposely ordered so that interviewers would be less likely to miscode race for members of minority populations.

For the 33 cases in which the adult respondent verified their race as Native Hawaiian, the value for the race item was not edited and was retained in the data delivery files. Cases in which verification was not obtained or if the case was finalized as non-working, unknown person, or no answer, the race item was set to missing and later imputed. The remaining cases were coded based on the responses from the recontact.

Table 5-1. Summary of final cases disposition

Final Disposition	Count
Hawaiian	33
White	62
Asian	6
American Indian/Alaska Native/White	4
Hispanic	6
Black	3
Refusal	5
Other (No such person)	6
NW	7
No Answer	5
Total sample	137

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

## REFERENCES

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

The California Health Interview Survey (CHIS) is a collaboration of:



UCLA Center for  
Health Policy Research



California Department  
of Health Services



Public Health  
Institute

California Health Interview Survey (CHIS)

UCLA Center for Health Policy Research

10911 Weyburn Avenue, Suite 300

Los Angeles, California 90024-2887

Phone 310-794-0925

Toll Free 1-866-275-2447

Fax 310-794-2686

[chis@ucla.edu](mailto:chis@ucla.edu)

[www.chis.ucla.edu](http://www.chis.ucla.edu)