November, 2011

CHIS 2009 Methodology Report Series

Report 3

Data Processing Procedures

CALIFORNIA HEALTH INTERVIEW SURVEY

<u>CHIS 2009 METHODOLOGY SERIES</u>

REPORT 3

DATA PROCESSING PROCEDURES

November, 2011

www.chis.ucla.edu

This report describes the data processing and editing procedures for CHIS 2009 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 geography and race and ethnicity survey items.

Suggested citation:

California Health Interview Survey. *CHIS 2009 Methodology Series: Report 3 – Data Processing Procedures*. Los Angeles, CA: UCLA Center for Health Policy Research, 2011.

Data Processing Procedures is the third report in a series of methodological reports describing the 2009 California Health Interview Survey (CHIS 2009). 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 Public Health, and the Department of Health Care Services. Westat was responsible for data collection and the preparation of five methodological reports from the 2009 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 2009 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 2009. 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 geographic coding. The next chapter describes how the race and ethnicity survey items were coded for CHIS.

TABLE OF CONTENTS

List of Tables

TABLE OF CONTENTS

List of Tables (continued)

1. CHIS 2009 SAMPLE DESIGN AND METHODOLOGY SUMMARY

1.1   Overview

The California Health Interview Survey (CHIS) is a population-based telephone survey of California's population conducted every other year since 2001. CHIS is the largest health survey conducted in any state and one of the largest health surveys in the nation. CHIS is based at the UCLA Center for Health Policy Research (CHPR) and is conducted in collaboration with the California Department of Public Health (CDPH) and the Department of Health Care Services (DHCS). 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 health related issues.

The sample is designed to meet and optimize two objectives:

- provide estimates for large- and medium-sized counties in the state, and for groups of the smallest counties (based on population size), and

- provide statewide estimates for California's overall population, its major racial and ethnic groups, as well as several Asian and Latino ethnic subgroups.

The 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 CHIS 2009, the fifth CHIS data collection cycle, which was conducted between September 2009 and April 2010. The previous CHIS cycles (2001, 2003, 2005, and 2007) are described in similar series, available at http://www.chis.ucla.edu/methods.html.

CHIS data and results are used extensively by federal and State agencies, local public health agencies and organizations, advocacy and community organizations, other local agencies, hospitals, community clinics, health plans, foundations, and researchers. The data are widely used for analyses and publications to assess public health and health care needs, to develop and advocate policies to meet those needs, and to plan and budget health care coverage and services.

1.2 Sample Design Objectives

To achieve the sample design objectives stated above, CHIS employed a multi-stage sample design. For the first time, the random-digit-dial (RDD) sample included telephone numbers assigned to both landline and cellular service. For the landline RDD sample, the state was divided into 44 geographic sampling strata, including 41 single-county strata and three multi-county strata comprised of the 17 remaining counties. Within each geographic stratum, residential telephone numbers were selected, and within each household, one adult (age 18 and over) respondent was randomly selected. In those households with adolescents (ages 12-17) and/or children (under age 12), one adolescent and one child were randomly selected; the adolescent was interviewed directly, and the adult most knowledgeable about the child's health completed the child interview.

Table 1-1 shows the 44 sampling strata, which include 41 independent county strata. A sufficient number of adult interviews were allocated to each stratum to support the first sample design objective—to provide health estimates for adults at the local level. The geographic stratification of the state was the same as that used since CHIS 2005. In the first two CHIS cycles there were 41 total sampling strata, including 33 individual counties. The CHIS 2009 samples in Humboldt, Marin, and San Diego Counties were enhanced with additional funding.

The main landline RDD CHIS sample size is sufficient to accomplish the second objective. To increase the precision of estimates for Koreans and Vietnamese, areas with relatively high concentrations of these groups were sampled at higher rates. These geographically targeted oversamples were supplemented by telephone numbers associated with group-specific surnames drawn from listed telephone directories to further increase the sample size for Koreans and Vietnamese. CHIS 2009 included additional Korean and Vietnamese oversamples conducted on behalf of the National Cancer Institute.

To help compensate for the increasing number of households without landline telephone service, a separate RDD sample was drawn of telephone numbers assigned to cellular service. In CHIS 2009, the goal was to complete approximately 2,500 interviews statewide with adults from the cell-phone sample . The CHIS 2009 cell-phone sampled from the CHIS 2007 cell-phone sample in two significant ways. First, all cell-phone sample cases were eligible for the extended interview regardless of the presence of a landline phone. The landline and cell samples, therefore, overlap and contrasts to CHIS 2007 when cell-phone cases with a landline telephone were screened out to limit the cell-phone sample to "cell-phone only" cases. This change was made due to the large and potentially unique characteristics of

telephone users who possess both a landline and cell-phone, but rely principally on their cell-phone for communication and would otherwise be excluded from the sample. The second change to the cell-phone sample was the inclusion of child and adolescent extended interviews. About 200 teen interviews and nearly 500 child interviews were completed from the cell-phone sample in CHIS 2009. Because data are not available for numbers assigned to cellular service to support the same level of geographic stratification as the landline sample, the cell RDD sample was stratified by area code. If the sampled number was shared by two or more adult members of a cell-only household, one household member was selected for the adult interview. Otherwise, the adult owner of the sampled number was selected.

Table 1-1.   California county and county group strata used in the CHIS 2009 sample design

1. Los Angeles	7. Alameda	27. Shasta
1.1 Antelope Valley	8. Sacramento	28. Yolo
1.2 San Fernando Valley	9. Contra Costa	29. El Dorado
1.3 San Gabriel Valley	10. Fresno	30. Imperial
1.4 Metro	11. San Francisco	31. Napa
1.5 West	12. Ventura	32. Kings
1.6 South	13. San Mateo	33. Madera
1.7 East	14. Kern	34. Monterey
1.8 South Bay	15. San Joaquin	35. Humboldt
2. San Diego	16. Sonoma	36. Nevada
2.1 N. Coastal	17. Stanislaus	37. Mendocino
2.2 N. Central	18. Santa Barbara	38. Sutter
2.3 Central	19. Solano	39. Yuba
2.4 South	20. Tulare	40. Lake
2.5 East	21. Santa Cruz	41. San Benito
2.6 N. Inland	22. Marin	42. Colusa, Glen, Tehama
3. Orange	23. San Luis Obispo	43. Plumas, Sierra, Siskiyou, Lassen,
4. Santa Clara	24. Placer	Modoc, Trinity, Del Norte
5. San Bernardino	25. Merced	44. Mariposa, Mono, Tuolumne,
6. Riverside	26. Butte	Alpine, Amador, Calaveras, Inyo

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

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 analysis of 2000 Census data to identify 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 2009 data collection under contract with the UCLA Center for Health Policy Research. For the landline RDD sample, Westat staff interviewed one randomly selected adult in each sampled household, and sampled one adolescent and one child if present in the household and the sampled adult was the parent or legal guardian. Up to three interviews could have been completed in each household. In households with children where the sampled adult was not the screener respondent, children and adolescents could be sampled as part of the screening interview, and the extended child (and adolescent) interviews could be completed before the adult interview. This "child-first" procedure was new for CHIS 2005 and has been continued in subsequent CHIS cycles; this procedure substantially increases the yield of child interviews. While numerous subsequent attempts were made to complete the adult interview, there were completed child and/or adolescent interviews in households for which an adult interview was not completed. Table 1-2 shows the number of completed adult, child, and adolescent interviews in CHIS 2009 by the type of sample (landline RDD, surname list, and cell RDD).

Table 1-2.   Number of completed CHIS 2009 interviews by type of sample and instrument

Type of sample	Adult	Child	Adolescent
Total all samples	47,614	8,945	3,379
Landline RDD	42,682	7,918	3,002
Surname list	1,885	545	178
Cell RDD	3,047	482	199

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Interviews in all languages were administered using Westat's computer-assisted telephone interviewing (CATI) system. The average adult interview took about 40 minutes to complete. The average child and adolescent interviews took about 16 minutes and 18 minutes, respectively. For "child-first" interviews, additional household information asked as part of the child interview averaged about 9 minutes. Interviews in non-English languages generally took longer to complete. More than 12 percent of the adult interviews were completed in a language other than English, as were almost 24 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).

1.4 Response Rates

The overall response rate for CHIS 2009 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 one or more selected persons to complete the extended interview). To maximize the response rate, especially at the screener stage, an advance letter in five languages was mailed to all landline sampled telephone numbers for which an address could be obtained from reverse directory services. An advance letter was mailed for approximately 58 percent of the landline RDD sample telephone numbers, and 82 percent of list sample numbers. Addresses were not available for the cell sample. As in CHIS 2005 and 2007, a $2 bill was included with the advance letter to promote cooperation.

The CHIS 2009 screener completion rate for the landline and samples was 36.1 percent, and was higher for households that were sent the advance letter. For the cell phone sample, the screener completion rate was 19.3 percent in all households. The extended interview completion rate for the landline sample varied across the adult (49.0 percent), child (72.9 percent) and adolescent (42.8 percent) interviews. The adolescent rate includes getting permission from a parent or guardian. The adult interview completion rate for the cell sample was 56.2 percent. Multiplying the screener and extended rates gives an overall response rate for each type of interview. The percentage of households completing one or more of the extended interviews (adult, child, and/or adolescent) is a useful summary of the overall performance of the landline sample. For CHIS 2009, the landline sample household response rate was 19.7 percent (the product of the screener response rate and the completion rate at the household level of 54.7 percent). All of the household and person level response rates vary by sampling stratum. For more information about the CHIS 2009 response rates, please see *CHIS 2009 Methodology Series: Report 4 – Response Rates*.

Table 1-3.   CHIS 2009 survey topic areas by instrument

Health status	Adult	Teen	Child
General health status, height and weight	✓	✓	✓
Days missed from school due to health problems	✓	✓	✓

Health conditions	Adult	Teen	Child
Asthma	✓	✓	✓
Diabetes, gestational diabetes, pre-diabetes/borderline	✓		
Heart disease, high blood pressure	✓		
Physical disability	✓		
Developmental assessment and developmental conditions			✓

Mental health	Adult	Teen	Child
Mental health status	✓	✓	✓
Perceived need, access and utilization of mental health services	✓	✓	✓
Suicide ideation and attempts	✓		

Health behaviors	Adult	Teen	Child
Dietary intake, fast food, high sugar diet	✓	✓	✓
Physical activity and exercise	✓	✓	✓
Walking for transportation and leisure	✓		
Sedentary time		✓	✓
Flu Shot	✓		✓
Alcohol and tobacco use	✓	✓	
Illegal drug use		✓	
Sexual behavior	✓	✓	
HIV/STI testing		✓	
Sun exposure	✓	✓	

Women's health	Adult	Teen	Child
Mammography screening, hormone replacement therapy	✓		
Age at menarche, live births, menopause, birth control medications	✓		
Pregnancy status	✓	✓	

Cancer history and prevention	Adult	Teen	Child
Family history	✓		
Colorectal cancer screening, prostate specific antigen (PSA) test	✓		

Dental health	Adult	Teen	Child
Last dental visit, main reason haven't visited dentist		✓	✓

Table 1-3.   CHIS 2009 survey topic areas by instrument (Continued)

Food environment	Adult	Teen	Child
Availability of food in household over past 12 months	✓		
Brought lunch to school from home		✓	
Doctor discussed nutrition/physical activity		✓	✓

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 and medical care)	✓	✓	✓
Medical home	✓	✓	✓
Communication problems with doctor	✓		
Long-term care	✓		

Health insurance	Adult	Teen	Child
Current insurance coverage, spouse's coverage, who pays for coverage	✓	✓	✓
Health plan enrollment, characteristics and plan assessment	✓	✓	✓
Employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months, reason for lack of insurance	✓	✓	✓
Medical debt, high deductible health plans	✓	✓	✓
Partial scope Medi-Cal, Medi-Cal deficit reduction act requirements	✓		

Public program eligibility	Adult	Teen	Child
Household poverty level	✓		
Program participation (TANF, CalWorks, Public Housing, Food Stamps, SSI, SSDI, WIC)	✓	✓	✓
Assets, alimony/child support/social security/pension	✓		
Medi-Cal and healthy families eligibility	✓	✓	✓
Reason for Medi-Cal non-participation among potential beneficiaries	✓	✓	✓

Neighborhood and housing	Adult	Teen	Child
Neighborhood safety, use of parks		✓	✓
Homeownership, length of time at current residence	✓		
Civic engagement		✓	✓
Social cohesion			✓

Emergency Preparedness	Adult	Teen	Child
Medication supply and basic preparedness	✓		

Interpersonal Violence	Adult	Teen	Child
Intrapersonal violence	✓		

Table 1-3.   CHIS 2009 survey topic areas by instrument (Continued)

Parental involvement/adult supervision	Adult	Teen	Child
Adult presence after school/knowledge of teen's activities, role models		✓	
Parental concerns/involvement			✓

Child care and school attendance	Adult	Teen	Child
Current child care arrangements			✓
Paid child care	✓		
First 5 California: Parent kit, educational TV programming			✓
Preschool/school attendance, name of school		✓	✓

Employment	Adult	Teen	Child
Employment status, spouse's employment status	✓		
Hours worked at all jobs	✓		

Income	Adult	Teen	Child
Respondent's and spouse's earnings last month before taxes	✓		
Household income (annual before taxes)	✓		
Number of persons supported by household income	✓		

Respondent characteristics	Adult	Teen	Child
Race and ethnicity, age, gender, height, weight, education	✓	✓	✓
Veteran status	✓		
Marital status, registered domestic partner status	✓		
Sexual orientation	✓	✓	
Language spoken with peers, language of TV, radio, newspaper used	✓		
Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home, English language proficiency	✓	✓	✓

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Historically, the CHIS response rates are comparable to response rates of other scientific telephone surveys in California, such as the California Behavioral Risk Factor Surveillance System (BRFSS) Survey. It has become increasingly difficult, however, to compare the CHIS and BRFSS response rates due to changes in the BRFSS response rate calculation methods. 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. The 2009 BRFSS, for example, shows the refusal rate for the California (32.2%) is the highest in the nation and more than twice the national median (15.7%).[1] Survey response rates tend to

---

[1] As reported in the Behavioral Risk Factor Surveillance System 2009 Summary Data Quality Report (Version #1 – Revised: 04/27/2010, available online at ftp://ftp.cdc.gov/pub/Data/Brfss/2009_Summary_Data_Quality_Report.pdf

be lower in California than nationally, and over the past decade response rates have been declining both nationally and in California. Further information about CHIS data quality and nonresponse bias is available at http://www.chis.ucla.edu/dataquality.html.

Adults who completed at least approximately 80 percent of the questionnaire (i.e., through Section K (on employment, income, poverty status, and food security), after all follow-up 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 those cases that did not complete the entire interview.

Proxy interviews were allowed for frail and ill persons over the age of 65 who were unable to complete the extended adult interview in order to avoid biases for health estimates of elderly persons that might otherwise result. Eligible selected persons were recontacted and offered a proxy option. For 283 elderly adults, a proxy interview was completed by either a spouse/partner or adult child. A reduced questionnaire, with questions identified as appropriate for a proxy respondent, was administered. (Note: questions not administered in proxy interviews are given a value of "-2" in the data files.)

1.5   Weighting the Sample

To produce population estimates from the CHIS data, weights are applied to the sample data to compensate for the probability of selection and a variety of other 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. The weighting procedures used for CHIS 2009 accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons;

- 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" (the inverse of the probability of selection of the telephone number) and a variety of adjustment factors. The household weight is used to compute a person-level weight, which includes 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 population control totals from an independent data source (see below). The procedure requires iteration to make sure all the control totals, or raking dimensions, are simultaneously satisfied within a specified tolerance.

Population control totals of the number of persons by age, race, and sex at the stratum level for CHIS 2009 were created primarily from the California Department of Finance's 2009 Population Estimates and 2009 Population Projections. The raking procedure used 11 raking dimensions, which are combinations of demographic variables (age, sex, race, and ethnicity), geographic variables (county, Service Planning Area in Los Angeles County, and Health Region in San Diego County), 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 differential response rates from households with and without landline telephones. One limitation of using 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 nine or more unrelated persons). These persons were excluded from the CHIS target population and as a result, the number of persons living in group quarters was estimated and removed from the Department of Finance control totals prior to raking.

1.6 Imputation Methods

Missing values in the CHIS data files were replaced through imputation for nearly every variable. This was a massive task designed to enhance the analytic utility of the files. Westat imputed missing values for those variables used in the weighting process and UCLA-CHPR staff imputed values for nearly all other variables.

Two different imputation procedures were used by Westat to fill in item nonresponse for items essential for weighting the data. The first imputation technique was a completely random selection from the observed distribution of respondents. This method was used only for a few variables when the percentage of the items missing was very small. The second technique was hot deck imputation without replacement. The hot deck approach is probably the most commonly used method for assigning values for

missing responses. 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 a survey 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 the same items in CHIS 2003, CHIS 2005, CHIS 2007, and CHIS 2009 (i.e., race, ethnicity, home ownership, and education).

UCLA-CHPR imputed missing values for nearly every variable in the data files other than those imputed by Westat and some sensitive variables in which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2009 were low, with most variables missing valid responses for less than 2% of the sample. However, there were a few exceptions where item nonresponse rate was greater than 25% such as household income.

The imputation process conducted by UCLA-CHPR started with data editing, sometimes referred to as logical or relational imputation: for any missing value, a valid replacement value was sought based on known values of other variables of the same respondent or other sample(s) from the same household. For the remaining missing values, model-based hot-deck imputation with donor replacement was used. This method replaces a missing value for one respondent using a valid response from another respondent with similar characteristics as defined by a generalized linear model with a set of control variables (predictors). The link function of the model is corresponding to the nature of the variable being imputed, e.g. generalized linear regression for continuous variables, logistic regression for binary and multinomial variables, and negative binomial regression for counts variables. The donors and recipients are grouped based on their predicted values from the model.

Control variables (predictors) used in the model to form donor pools for hot-decking always included the following: gender, age group, race/ethnicity, poverty level (based on household income), educational attainment, and region. Other control variables were also used depending on the nature of the imputed variable. Among the control variables, gender, age, race/ethnicity and regions were imputed by Westat. UCLA-CHPR then imputed household income and educational attainment in order to impute other variables. Household income, for example, was imputed using the hot-deck method within ranges from a set of auxiliary variables such as income range and/or poverty level.

The imputation order of the other variables followed the questionnaire. After all imputation procedures were complete, every step in the data quality control process is performed once again to ensure consistency between the imputed and nonimputed values on a case-by-case basis.

1.7   Methodology Report Series

A series of five methodology reports is available with more detail about the methods used in CHIS 2009:

- 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 http://www.chis.ucla.edu or contact CHIS at CHIS@ucla.edu.

2. DATA EDITING PROCEDURES

Survey data for all CHIS 2009 samples – landline, list, and cell– were collected using the same computer-assisted telephone interview (CATI) system. While the screening interview varied somewhat by sample, the same editing procedures were followed for all CHIS 2009 cases.

In a CATI environment, the data collection and interview process is controlled using a series of computer programs to ensure consistency and quality. (*CHIS 2009 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 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 for post-interview editing and allows most 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 2009 survey data were carefully examined and edited before delivering final data files to UCLA. 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 2009 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 (i.e., "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 in turn.

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 an electronic problem sheet for the case. Problem sheets from all the TRCs were transmitted to a single staff member who distributed them via Westat's corporate intranet 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. An actual hard-copy form was not necessary as the web form was reviewed on processed on-line.

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 E, 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 back up properly"). The solution for these types of cases was to re-field the interview with a message stating, "Case will restart in Section E. Re-ask beginning with screen AD1." Note that questions from earlier CHIS cycles that were also asked in CHIS 2009 retained their original CATI screen names, in addition to having a sequential number appropriate to the 2009 interview. In this example, the first question in Section E for CHIS 2009 has screen name AD1.

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 a change in the person named as most knowledgeable about the sampled child was needed. 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 in special entry windows in CATI by interviewers when they want to record respondents' statements but are unable to enter as a standard response in the instrument. For CHIS 2009, some phrases were merely an elaboration of a previously recorded response, an expression of opinion, or comments unrelated to the survey which did not necessarily require modifying or updating survey responses. Other times, comments were substantive to data quality and indicated that an update was needed. For example, if an answer for child's weight was outside the existing range of child's weight range programmed in CATI, the interviewer first asks respondent to confirm, enters DK as the answer, and adds a comment stating actual weight.

Comments were also used to identify specific responses that could not be coded using the existing response option set. In previous CHIS cycles, response option sets for some question items were amended or updated in the CATI system during the survey field period. Other such changes have occurred in preparation for the next CHIS cycle. These changes have helped reduce the number of interviewer comments and lessen the amount of data preparation work. For CHIS 2009, the only changes

to the response option sets were made after data collection had been completed. New codes created after a number of similar of responses are found during the review of "other specify" text. The decision to create a new response option is made if the total number of entries that can be grouped under a new category is larger than the number of entries for any existing codes. Table 2-1 provides examples of items and responses that interviewers initially had difficulty coding.

Weekly meetings between data preparation and project staff during data collection covered data-related issues, reviewed comments, and established case-specific procedures for handling pending or interim problem cases. Comments and cases under review included both complete and incomplete (interim status) interviews.

Table 2-1.   Examples of difficult responses to code in CHIS 2009

CATI Screen ID	Question and response options	Respondents' answers

AH49	Is your MediCARE coverage provided through an HMO?
	
	   1. YES
	   2. NO
	   -7. REFUSED
	   -8. DON'T KNOW
		"I get coverage through my union."

AI22A	What is the name of your {Medi-Cal} health plan?
	   1. (List County Plans),
	   -7.REFUSED
	   -8. DON'T KNOW
		"Kaiser"

2.3   Coding with Text Strings

Most items in the CHIS 2009 had only closed-ended response options, so coding of open-ended responses was not needed. The survey had a number of other-specify questions, however, that required coding of narrative text strings recorded by interviewers. Other-specify questions 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 2009 extended interviews included:

- Racial/ethnic ancestry (AA5, AA5A, AA5E, AA5E1);
- Tribal names (AA5B, AA5D);

- Sexual orientation (AD46);
- Country of birth (AH33, AH34, AH35);
- Languages spoken at home (AH36);
- Place visited for health care (AH3);
- Place visited for flu vaccine (AB57);
- Sexual orientation (AD46);
- How first found out about breast cancer (AB60);
- Health insurance coverage items (AI15, KAI15, AI15A, KAI15A, AI17A, KAI17A, AI45, KAI45, AI45A, KAI45A, AI36, KAI36, AI24, KAI24, AL19);
- Child/adolescent health insurance coverage items (CF7, KCF7, CF18, KCF18, IA18, KIA18, CF29, KCF29, IA29, KIA29, CF1A, CF2A, KCF2A, IA2A, IA7,KIA7).
- Adult/child/adolescent Insurance plan names (AH50, AI22A, MA2, MA7, KAH50, KAI22A, KMA2, KMA7);
- Reasons no longer receiving treatment for mental or emotional health or for an alcohol or drug problem (AF80);
- Country of birth (AI56, AI56C, AI56T);
- Languages used by doctor to speak to respondent (AJ50);
- Relationship to respondent of perpetrators of violence (AJ69);
- Child condition or disability (CA10A);
- Adolescent race and ethnicity (TI1A, TI2, TI2A,TI2C,TI2D,TI2D1);
- Child race and ethnicity (CH2, CH3, CH4, CH6, CH7, CH7A);
- Child/teen languages spoken at home (CH17, TI7);
- Child/mother/father place of birth (CH8, CH11, CH14);
- Adolescent country of birth (TI3);
- Child/adolescent school name (CB22, TA4B);
- Child/adolescent usual source of health care (CD3, TF2);
- Place child got last flu vaccine (CD42);
- Type of STD adolescent tested for (TH32);
- Reason for adolescent not visiting dentist in past year (TM1);
- Race/ethnicity respondent identifies with (DM/DMA1); and
- Reasons for unfair treatment (DMB9A_OV, DMC6A_OV)[2].
- Person teen admires (TH23)
- Race or ethnicity of person teen admires (TH25)

---

[2] The data manager did not review or update discrimination module (DM) data unless a specific correction was requested by interviewer, a very rare occurrence.

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 2-2.

CHIS 2009 did not collect any 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 2-2.   Response codes added to CHIS 2009

Questionnaire version	Variable Name	Question Name	New code	Response description
Adult	AD24	**QA09_E21**	5	Sonogram
Adult	AM35	**QA09_M7**	5	Not Applicable
Adult	AI22A	**QA09_H57**	92	Other-Specify

Source: UCLA Center for Health Policy Research, 2009 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 re-ask the question to be certain the respondent understood what was being asked as needed. Hard-range edits prevented recording unacceptable values. For example, for a question on how many glasses of juice the adolescent respondent had the previous day, the soft range is 0-9, the hard range 0-20. 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, 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; approximately 120,000 database values were updated and verified for CHIS 2009. The majority of updated values resulted from the addition of insurance plan names for the adult, adolescent, and child extended interviews. These were updated to existing codes, and different spellings of the same name were consolidated.

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 checked 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 were interrupted unexpectedly by a power failure or system crash. The precise number of CHIS 2009 cases restored using audit trail information is not known, but recovery was needed for a limited number of cases due to power failure during inclement weather.

2.5 Data Conversion and Delivery

The CATI survey data were collected and stored in a hierarchical database to improve data efficiency and enhance performance while interviewing. For delivery to UCLA, these data were converted to SAS datasets using Westat's CATISAS, a series of SAS macros programs. CATISAS extracts data attributes 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. The SAS data sets created by the CATISAS macro were then combined and further processed with SAS to create the delivery files.

During the conversion process from CATI database elements to SAS files, diagnostic edit checks were run on the database. Frequencies for categorical data were run and examined. These reviews were made to ensure that errors had not been inadvertently introduced into 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;

- Weights (final combined and separate sample weight and replicate weights);

- Imputation flags, and Geocoding results for all completed households

3. GEOGRAPHIC CODING

For CHIS 2009, Westat was responsible for delivering coded survey data for items from the adult extended interview, or the child interview in "child-first" cases, related to geographic location of residence. The self-reported county was used to assign cases to landline sample strata as described in *CHIS 2009 Methodology Series: Report 1: Sample Design.* Westat also prepared and delivered more specific geocodes based on the respondent-reported address and other information. The geographic coding process for CHIS 2009 used the 2009 NAVTEQ database of roads and corresponding NAVTEQ Census Block boundary definitions (NAVTEQ North America, LLC, http://www.navteq.com/).

3.1   County of Residence

The CHIS 2009 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[3], interviewers verified the street address and Zip code with the adult respondent (AO1) and then collected the name of a nearby cross-street (AM9). These same questions were asked of adults who completed the child interview under the "child first" protocol. The child-first protocol allowed completion of the child interview before the adult extended interview was conducted. (See *CHIS 2009 Methodology Series: Report 2 – Data Collection Methods* for details regarding the child-first protocol.)

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.

Because telephone numbers were assigned to sampling strata based on the telephone area code and exchange (see *CHIS 2009 Methodology Series: Report 1 - Sample Design*, 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 interview to the correct stratum that the adult self-reported as the residence.

---

[3] 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.

The following two questions from the adult extended interview were used to make the self-reported stratum assignment during data collection[4]:

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

and

AM7.   "What is your Zip code?"

Table 3-2 at the end of this chapter shows a listing of Zip codes within each landline sampling stratum[5] for CHIS 2009. The final self-reported stratum for the analysis file was determined by applying the geocodes developed by UCLA and CHIS staff. See *CHIS 2009 Methodology Series: Report 5 - Weighting and Variance Estimation*, Section 8.5, for a fuller discussion of this process.

The final distribution of completed landline sample adult extended interview cases by self-reported and original sampling stratum is presented in Table 3-3 at the end of this chapter. 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 of residence.

3.2   Geocoding Process

Two methods of geocoding using NAVTEQ software were employed for CHIS 2009. The first option was to have the software automatically match (batch match) the input addresses to a spatial database of roads, which returned the address's latitude/longitude, state fips and county fips. If the software was unable to match to the street address, it automatically matched to the geographic zip centroid as a fallback. In such cases, the latitude/longitude, state fips and county fips of the zip centroid was provided.

The second method performed the same batch process as described in option 1 above, but did not automatically default to a zip centroid match. If a batch match was not obtained, Westat interactively examined the unmatched records (excluding PO boxes and rural routes) to try and determine the reason why the software could not automatically match the address. Sometimes this was due to misspelled street names, city names, etc., or to missing house numbers. Westat corrected the address to

---

[4] The same questions were asked in the child-first interview, as KAH42 and KAM7.

[5] The cell sample used larger geographic areas as strata. See *CHIS 2009 Methodology Series:Report 1 – Sample Design.*

match the street database, or matched to the segment's nearest intersection. If the street address or nearest intersection could not be identified, Westat would then match to geographic zip centroid. The frequencies of assigned geocodes by rule and sample type are shown in table 3-1.

Table 3-1.   Number of Geocodes Assigned by Rule and by Sample Type

Rule	Sample			
	Landline	Cell	List	Total
1 - A01 Address	33,818	2,597	1,836	38,251
2 - A02 & AM7 Address fields	0	0	0	0
3 - EXTD (incentive) Address fields	0	69	0	69
4 - MAIL Address fields	1,560	0	78	1,638
5 - Cross street (AM8 and AM9) & ZIP (AM7)	1,806	71	53	1,930
7 - AM7 - Zip only (zip centroid)	2,507	200	100	2,807
9 - BESTZIP1-6 - Zip centroid, using BESTZIP1-6	1	2	0	3
10 - Geocoded to the imputed GEO_TRACT level	5	78	0	83
11 - Geocoded to the phone lookup GEO_FIPCNTY level	0	30	0	30
TOTAL	44,697	3,047	2,067	49,811

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey

Table 3-2.   Zip codes within landline 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, 90099, 90101, 90102, 90103, 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, 90631, 90633, 90637, 90638, 90639, 90640, 90650, 90651, 90652, 90659, 90660, 90661, 90662, 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, 90809, 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, 91182, 91184, 91185, 91188, 91189, 91191, 91199, 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, 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, 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, 93563, 93584, 93586, 93590, 93591, 93599

Table 3-2.   Zip codes within landline 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, 92009, 92008, 92009, 92010, 92011, 92013, 92014, 92018, 92019, 92020, 92021, 92022, 92023, 92024, 92025, 92026, 92027, 92028, 92029, 92030, 92033, 92036, 92038, 92039, 92040, 92046, 92049, 92051, 92052, 92054, 92055, 92056, 92057, 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, 92259, 92536, 92672
3	Orange	90620, 90621, 90622, 90623, 90624, 90630, 90632, 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, 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, 94301, 94302, 94303, 94304, 94305, 94306, 94309, 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, 95103, 95106, 95108, 95109, 95110, 95111, 95112, 95113, 95115, 95116, 95117, 95118, 95119, 95120, 95121, 95122, 95123, 95124, 95125, 95126, 95127, 95128, 95129, 95130, 95131, 95132, 95133, 95134, 95135, 95136, 95138, 95139, 95140, 95141, 95148, 95150, 95151, 95152, 95153, 95154, 95155, 95156, 95157, 95158, 95159, 95160, 95161, 95164, 95170, 95172, 95173, 95190, 95191, 95192, 95193, 95194, 95196,
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, 92331, 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, 93562, 93592

Table 3-2.   Zip codes within landline sampling stratum (continued)

	Stratum	Zip code
6	Riverside	91752, 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, 92501, 92502, 92503, 92504, 92505, 92506, 92507, 92508, 92509, 92513, 92514, 92515, 92516, 92517, 92518, 92519, 92521, 92522, 92530, 92531, 92532, 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	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, 94649, 94659, 94660, 94661, 94662, 94666, 94701, 94702, 94703, 94704, 94705, 94706, 94707, 94708, 94709, 94710, 94712, 94720
8	Sacramento	94203, 94204, 94205, 94206, 94207, 94208, 94209, 94211, 94229, 94230, 94232, 94234, 94235, 94236, 94237, 94239, 94240, 94244, 94245, 94246, 94247, 94248, 94249, 94250, 94252, 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, 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, 95860, 95864, 95865, 95866, 95867, 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, 94582, 94583, 94595, 94596, 94597, 94598, 94801, 94802, 94803, 94804, 94805, 94806, 94807, 94808, 94820, 94850
10	Fresno	93210, 93234, 93242, 93602, 93605, 93606, 93607, 93608, 93609, 93611, 93612, 93613, 93616, 93619, 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, 93723, 93724, 93725, 93726, 93727, 93728, 93729, 93730, 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

Table 3-2.   Zip codes within landline sampling stratum (continued)

	Stratum	Zip code
12	Ventura	91319, 91320, 91358, 91359, 91360, 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, 93036, 93040, 93041, 93042, 93043, 93044, 93060, 93061, 93062, 93063, 93064, 93065, 93066, 93093, 93094, 93099
13	San Mateo	94002, 94005, 94010, 94011, 94013, 94014, 94015, 94016, 94017, 94018, 94019, 94020, 94021, 94025, 94026, 94027, 94028, 94030, 94037, 94038, 94044, 94060, 94061, 94062, 94063, 94064, 94065, 94066, 94070, 94074, 94080, 94083, 94096, 94098, 94128, 94401, 94402, 94403, 94404, 94497
14	Kern	93203, 93205, 93206, 93215, 93216, 93220, 93222, 93224, 93225, 93226, 93238, 93240, 93241, 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, 93554, 93555, 93556, 93558, 93560, 93561, 93581, 93596
15	San Joaquin	95201, 95202, 95203, 95204, 95205, 95206, 95207, 95208, 95209, 95210, 95211, 95212, 95213, 95215, 95219, 95220, 95227, 95231, 95234, 95236, 95237, 95240, 95241, 95242, 95253, 95258, 95267, 95269, 95296, 95297, 95304, 95320, 95330, 95336, 95337, 95366, 95376, 95377, 95378, 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	95230, 95307, 95313, 95316, 95319, 95323, 95326, 95328, 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, 93254, 93427, 93429, 93434, 93436, 93437, 93438, 93440, 93441, 93454, 93455, 93456, 93457, 93458, 93460, 93463, 93464
19	Solano	94510, 94512, 94533, 94534, 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

Table 3-2.   Zip codes within landline sampling stratum (continued)

Stratum	Zip code
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, 93475, 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, 95343, 95344, 95348, 95365, 95369, 95374, 95388
26 Butte	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
27 Shasta	96001, 96002, 96003, 96007, 96008, 96011, 96013, 96016, 96017, 96019, 96028, 96033, 96040, 96047, 96049, 96051, 96062, 96065, 96069, 96070, 96071, 96073, 96079, 96084, 96087, 96088, 96089, 96095, 96096, 96099
28 Yolo	95605, 95606, 95607, 95612, 95616, 95617, 95618, 95627, 95637, 95653, 95679, 95691, 95694, 95695, 95697, 95698, 95776, 95798, 95799, 95937
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	92222, 92227, 92231, 92232, 92233, 92243, 92244, 92249, 92250, 92251, 92257, 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, 93636, 93637, 93638, 93639, 93643, 93644, 93645, 93653, 93669
34 Monterey	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, 95039
35 Humboldt	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 Nevada	95712, 95724, 95728, 95924, 95945, 95946, 95949, 95959, 95960, 95975, 95977, 95986, 96111, 96160, 96161, 96162
37 Mendocino	95410, 95415, 95417, 95418, 95420, 95427, 95428, 95429, 95432, 95437, 95445, 95449, 95454, 95456, 95459, 95460, 95463, 95466, 95468, 95469, 95470, 95481, 95482, 95488, 95490, 95494, 95585, 95587
38 Sutter	95645, 95659, 95668, 95674, 95676, 95953, 95957, 95982, 95991, 95992, 95993
39 Yuba	95692, 95901, 95903, 95918, 95919, 95922, 95935, 95961, 95962, 95972, 95981
40 Lake	95422, 95423, 95424, 95426, 95435, 95443, 95451, 95453, 95457, 95458, 95461, 95464, 95467, 95485, 95493

Table 3-2.   Zip codes within landline sampling stratum (continued)

	Stratum	Zip code
41	San Benito	95023, 95024, 95043, 95045, 95075
42	Colusa, Glenn, Tehama,	95912, 95913, 95920, 95932, 95939, 95943, 95950, 95951, 95955, 95963, 95970, 95979, 95987, 95988, 96021, 96022, 96029, 96035, 96055, 96059, 96061, 96063, 96074, 96075, 96076, 96078, 96080, 96090, 96092
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	95527, 95531, 95532, 95538, 95543, 95548, 95552, 95563, 95567, 95568, 95595, 95910, 95915, 95923, 95934, 95936, 95944, 95947, 95956, 95971, 95980, 95983, 95984, 96006, 96009, 96010, 96014, 96015, 96020, 96023, 96024, 96025, 96027, 96031, 96032, 96034, 96037, 96038, 96039, 96041, 96044, 96046, 96048, 96050, 96052, 96054, 96056, 96057, 96058, 96064, 96067, 96068, 96085, 96086, 96091, 96093, 96094, 96097, 96101, 96103, 96104, 96105, 96106, 96108, 96109, 96110, 96112, 96113, 96114, 96115, 96116, 96117, 96118, 96119, 96121, 96122, 96123, 96124, 96125, 96126, 96127, 96128, 96129, 96130, 96132, 96134, 96135, 96136, 96137
44	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, 95329, 95335, 95338, 95345, 95346, 95347, 95364, 95370, 95372, 95373, 95375, 95379, 95383, 95389, 95601, 95640, 95642, 95644, 95646, 95654, 95665, 95666, 95669, 95675, 95685, 95689, 95699, 96107, 96120, 96133

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Table 3-3.   Final distribution of adult extended completed cases by self-reported and original sampling stratum, landline/list sample for CHIS 2009

Stratum name	Sampling stratum count	Removed	Added	Final self-reported stratum count
1 - LOS ANGELES	8,710	523	544	8,731
2 - SAN DIEGO	5,014	16	14	5,012
3 - ORANGE	2,636	78	29	2,587
4 - SANTA CLARA	1,607	8	60	1,659
5 - SAN BERNARDINO	1,460	29	25	1,456
6 - RIVERSIDE	1,609	13	35	1,631
7 - ALAMEDA	1,191	44	9	1,156
8 - SACRAMENTO	1,226	8	12	1,230
9 - CONTRA COSTA	895	1	52	946
10 - FRESNO	667	12	8	663
11 - SAN FRANCISCO	757	21	9	745
12 - VENTURA	898	5	27	920
13 - SAN MATEO	590	45	16	561
14 - KERN	577	6	5	576
15 - SAN JOAQUIN	517	2	-	515
16 - SONOMA	512	4	15	523
17 - STANISLAUS	474	23	1	452
18 - SANTA BARBARA	613	13	2	602
19 - SOLANO	480	15	13	478
20 - TULARE	473	3	10	480
21 - SANTA CRUZ	503	29	10	484
22 - MARIN	2,048	15	4	2,037
23 - SAN LUIS OBISPO	478	7	13	484
24 - PLACER	502	32	31	501
25 - MERCED	493	5	24	512
26 - BUTTE	493	2	17	508
27 - SHASTA	502	6	30	526
28 - YOLO	524	20	7	511
29 - EL DORADO	494	3	23	514
30 - IMPERIAL	539	7	-	532
31 - NAPA	485	8	18	495
32 - KINGS	478	3	2	477
33 - MADERA	535	13	8	530
34 - MONTEREY	426	4	68	490
35 - HUMBOLDT	844	16	18	846
36 - NEVADA	537	18	9	528
37 - MENDOCINO	600	21	1	580
38 - SUTTER	468	15	22	475
39 - YUBA	466	49	9	426
40 - LAKE	525	6	1	520
41 - SAN BENITO	548	47	1	502
42 - TEHAMA, ETC.	382	27	1	356
43 - DEL NORTE, ETC.	403	4	23	422
44 - TUOLUMNE, ETC.	388	9	9	388
Total	44,567	1,235	1,235	44,567

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

3.3   School Name and Geographic Coding

In CHIS 2009, the child and adolescent interviews included an item that collected the name of the school attended by the selected child or adolescent (CB22 and TA4B, respectively). The sampled adult or the most knowledgeable adult (MKA) reported the child's school name, and the sampled adolescent answered for him- or herself. Interviewers recorded the respondent's answers as a verbatim text entry in the CATI system. School latitude and longitude were then assigned to each school-aged child and adolescent case for which a school name was reported.

A review of the child interview data showed a number of spelling problems associated with item CB22 ("What is the name of the school {CHILD NAME /AGE/SEX} goes to or last attended"?). In many problem cases, the English-speaking adult respondent was reporting a Spanish school name (and was speaking to an English speaking interviewer). Asian and some Latino respondents, whose first language is not English, had similar difficulties in accurately reporting or spelling the school name.

Westat data preparation staff used the California Department of Education California School Directory (http://www.cde.ca.gov/re/sd/), AltaVista http://www.altavista.com/web/map, and Public School Review (http://www.publicschoolreview.com/) in conjunction with the respondent's Zip code as resources to improve the quality of school names and their location before release to UCLA for geocoding. SAS statistical programming was used to merge in open text from CB22 and TA4B as well as county of residence with relevant data fields in the school list database. Full matches were assigned a successful matching code. For cases that could not be automatically matched using statistical programming (e.g. spelling errors, county mismatch), additional CHIS variables were used to accurately identify and manually assign the name of the school. These included age of respondent, Zip code, city, and county of home residence. Additional information in the state school database was used to verify the child or adolescent's school, including school district, school county, school city, school Zip code, and school grade range. Web-based searches were also used to assign geographic school information not found in the California School Directory.

For all matched public schools, latitude and longitude were provided in the state-issued school database of California. Geocoding for private schools was performed by UCLA. Cases for which the child or adolescent attended a home school or non-traditional program or where a school could not be identified were assigned an undetermined value. Children under the age of 5 years were assigned an inapplicable value.

4. RACE AND ETHNICITY CODING

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 specify" category as a text string. The procedures for coding these "other specify" responses into existing codes (up-coding) or leaving them in the other category are presented here.

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 Latino or Hispanic. If the response to this item was "yes," then a question (AA5) was asked about the specific origin (Mexican, etc.) and this includes an "other" category with responses entered by interviewers as text in question AA5OS. Item AA5A from CHIS 2007 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?" The race question allowed the respondent to indicate that 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 also to say "other" race. The "other specify" race was recorded in text (AA5AOS). Another item followed 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 2009 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 prior CHIS iterations. The methods used in the 2000 Census are available at the U.S. Census Bureau Website (http://www.census.gov/prod/www/abs/pl94-171.pdf) and documented in *Census 2000 Redistricting Data (Public Law 94-171) Summary File – Technical Documentation* (U.S. Census Bureau, 2001). 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 administrations 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 "American" 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 specific Hispanic 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.*