california health interview survey

November, 2011

CHIS 2009 Methodology Report Series

Report 5

Weighting and Variance Estimation

CALIFORNIA HEALTH INTERVIEW SURVEY

CHIS 2009 METHODOLOGY SERIES

REPORT 5

WEIGHTING AND VARIANCE ESTIMATION

November 2011

This report was prepared for the California Health Interview Survey by Ismael Flores Cervantes, Greg Norman, and J. Michael Brick of Westat.

www.chis.ucla.edu

This report describes the weighting and variance estimation methods used in CHIS 2009. This report presents the steps used to create the analytical weights for analyzing the data from the adult, child, and adolescent interviews.

Suggested citation:

California Health Interview Survey. *CHIS 2009 Methodology Series: Report 5 – Weighting and Variance Estimation*. Los Angeles, CA: UCLA Center for Health Policy Research, 2011.

The California Health Interview Survey is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Public Health, and the Department of Health Care Services. Funding for CHIS 2009 came from multiple sources: the California Department of Public Health, the Department of Health Care Services, the California Endowment, the National Cancer Institute, NIH Office of Behavioral and Social Sciences Research, Agency for Healthcare Research and Quality, the California Wellness Foundation, First 5 California, the California Department of Mental Health, the California Office of the Patient Advocate, Kaiser Permanente, Blue Shield of California Foundation, the California HealthCare Foundation, the San Diego County Human and Health Services Agency, Marin County Department of Health and Human Services, and the Humboldt County Department of Health and Human Services.

PREFACE

Weighting and Variance Estimation is the fifth in a series of methodological reports describing the 2009 California Health Interview Survey. The other reports are listed below.

CHIS is a collaborative project of the University of California, Los Angeles (UCLA) Center for Health Policy Research, the California Department of Public Health, the Department of Health Care Services, and the Public Health Institute. 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 weighting and variance estimation methods from CHIS 2009. The purpose of weighting the survey data is to permit analysts to produce estimates of the health characteristics for the entire California population and subgroups including counties, and in some cases, cities. This report presents the steps used to create the analytical weights for analyzing the data from the adult, child, and adolescent interviews.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

List of Appendices

TABLE OF CONTENTS (continued)

List of Appendices (Continued)

List of Tables

TABLE OF CONTENTS (Continued)

List of Tables (Continued)

List of Figures

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, Modoc, Trinity, Del Norte
4. Santa Clara	24. Placer	
5. San Bernardino	25. Merced	44. Mariposa, Mono, Tuolumne, Alpine, Amador, Calaveras, Inyo
6. Riverside	26. Butte	

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 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;

[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

- 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. WEIGHTING ADJUSTMENTS CHIS 2009 SAMPLE WEIGHTS

This chapter introduces the concept of weighting and provides some background on the weights developed for analyzing CHIS 2009 survey data. Weighting is a process that attempts to make estimates from survey respondents representative of the total population from which they were sampled by accounting for the chances of selecting units into the sample and making adjustments for imperfections in the frame and the data collection process. The process begins with a base weight that is then adjusted to account for additional stages of sampling, nonresponse and undercoverage.

As described in *CHIS 2009 Methodology Series: Report 1 - Sample Design*, CHIS 2009 has samples from four different kinds of frames; landline RDD (including geographically-based supplements), surname list, and cellular. One set of weights was produced for data analysis, combining these samples.

Although this chapter deals with the weights and their adjustments, it begins with the general reasons why fully adjusted weights should be used. It also describes the details, advantages, and disadvantages of weighting.

2.1 Weighting Approach

The approach used in CHIS 2009 weighting is a standard design-based, multiple-frame methodology that is consistent with the sampling methods used. The multiple-frame approach was used in CHIS 2007 to combine and weight the landline, surname list and cell phone sample in 2007, and the landline and surname samples in previous cycles of CHIS.

The procedures used in CHIS are consistent for all users and analyses. Using the same analytic methods in a unified procedure also makes it much simpler for analysts to examine characteristics for many issues, such as preparing estimates from the main and landline supplemental samples for San Diego, Humboldt, and Marin[2]. Operationally, the weighting steps are similar and can be applied at the same time across samples (whenever appropriate), streamlining the weighting process and reducing the time required to produce the weights.

[2] These are landline geographic supplemental samples.

Weights are applied to CHIS 2009 sample data to estimate aggregate statistics at the state and county levels. In particular, sample weighting was carried out to accomplish the following objectives:

- Compensate for differential probabilities of selection and sampling rates for households and persons;
- Reduce biases occurring because nonrespondents may have different characteristics from 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.

The combined landline/list/cell telephone sample weights were created to produce estimates that avoid the coverage bias of a landline sample that excludes cell-only households. These estimates should be more accurate for the entire California population than those from the landline/list sample only. As in previous cycles of CHIS a single weight was created for each adult, child, and adolescent completed interview in the samples.

2.2 Weighting Adjustments

The final weight for a completed CHIS interview is the product of a series of sequential adjustments. The starting point within each sampling stratum is the development of a base weight (Section 3.1) defined as the inverse of the probability of selection from the stratum frame. After creating the landline and surname list base weights (Section 3.1.1) and the cell phone base weights (Section 3.1.2) the base weights are adjusted for

- Remaining ported cell numbers not dialed (section 3.2);
- Sampled telephone numbers never dialed (section 3.3);
- Residual landline telephone numbers without full refusal conversion (section 3.4);
- Unknown residential status (Section 3.5);
- Supplemental list sample eligibility (Section 3.6);
- Screener interview nonresponse (Section 3.7); and

- Multiple telephone numbers and duplicate respondent adjustments (Section 3.8).

These adjustments are described in Chapter 3.

The household weight is then adjusted to create a person weight for each type of extended interview. For the adult weights, the following factors are included:

- Probability of selection of the adult (Section 4.1);
- Extended adult interview nonresponse adjustment (Section 4.2);
- Telephone type adjustment (Section 4.3);
- Composite weight adjustment for combining the landline and cell phone samples (Section 4.3); and
- Trimming (Section 4.4) and raking (Section 4.5) adjustment to person-level control totals.

The child and adolescent weights are more complex because of the method used to sample children (see *CHIS 2009 Methodology Series: Report 1 - Sample Design*). For these weights, the adjustment factors include:

- Section G adult extended interview nonresponse adjustment for households in which adolescents and children are sampled at the end of Section G of the adult interview (Section 5.1);
- Probability of selection of the child or adolescent (Section 5.2 and 6.1);
- Extended child and adolescent interview nonresponse adjustment (Section 5.3 and 6.2);
- Telephone type adjustment (Section 5.3 and 6.2);
- Composite weight adjustment for combining the landline and cell phone samples (Section 5.3 and 6.2); and
- Trimming (Section 5.3 and 6.2) and raking (Section 5.3 and 6.2) adjustment to person-level control totals.

The expressions for the weighting factors and adjustments for the person weights are given in Chapters 4, 5, and 6. The derivation of the population control totals is described in Chapter 7. The imputation process and the variables imputed to support the weighting process are described in Chapter 8. Chapter 9 discusses methods for variance estimation for CHIS 2009.

Appendix A contains tables showing the frame and sample sizes. Appendix B contains tables that show the effect of each step of the weighting process at the household and person levels. Throughout this report, we refer to specific tables and rows in Appendix B that indicate how the weights were adjusted.

2.3 Nonresponse adjustments

In an ideal survey, all the units in the inference population are in the sample frame and all those in the sample participate in the survey. In practice, neither of these conditions occurs. Some units are not included in the frame (undercoverage) and some of the sampled units do not respond (nonresponse). If undercoverage and nonresponse are not addressed, then estimates from the survey may be biased. In CHIS 2009, the weights of those who respond are adjusted to represent undercovered persons in the population and nonrespondents in the sample. The approaches used to account for these two sources of missing data begin with adjusting for nonresponse.

Nonresponse results in biases in survey estimates when the characteristics of respondents differ from those of nonrespondents. The size of the bias depends on the magnitude of this difference and the response rate (see Groves, 1989). The purpose of adjusting for nonresponse is to reduce the bias. A weighting class adjustment (see Brick and Kalton, 1996) method is the type of nonresponse adjustment procedure used in CHIS 2009. In this procedure, nonresponse adjustment weights are computed and applied separately by cell, where a cell is defined using characteristics known for both nonrespondents and respondents. For example, the county associated with each telephone number is known, even if there are some misclassifications in the assignment. Thus, county can be used to define cells, and weighting adjustments can be computed separately for each of these cells. The more similar either response patterns or survey characteristics are within the cells, the larger the bias reduction in the adjustment.

The drawback to nonresponse adjustment is that it increases the variability of the weights and increases the sampling variance of the estimates (Kish, 1992). A nonresponse adjustment is beneficial only when the reduction in bias more than compensates for the increase in variance. When the cells contain sufficient cases and the adjustment factors do not become inordinately large, the effect on variances is often modest. Large adjustment factors usually occur in cells with small numbers of respondents. To avoid this situation, cells with few cases are "collapsed" or combined to form a new cell with a larger number of cases.

The operational objective for nonresponse adjustment in CHIS 2009 was to define adjustment cells for which response rates vary considerably and to avoid cells with either a small number of cases or a large adjustment factor. Since county-level estimates are important, the county was nearly always included in the definition of the cells. Oh and Scheuren (1983) discuss some of the statistical features associated with making these adjustments.

As noted above, nonresponse adjustment classes can be formed only if data are available for both responding and nonresponding units. Since the nonresponse adjustment is done for each stage of data collection, the data available for forming cells are different for each stage. For screening interviews, the nonresponse unit is a household (or more accurately a telephone number), and data must be available for all households. For extended interviews, the nonresponse adjustment is done by type of person (adult, child, or adolescent). At this level, data from the screening interview can be used to define cells.

The approach to adjusting for undercoverage is somewhat different from that for nonresponse because noncovered units or persons were never eligible to be sampled. The undercoverage adjustment procedure uses data from external sources (control totals) in a process called poststratification (Holt and Smith, 1979). The primary objective of poststratification is to dampen potential biases arising from a combination of response errors, sampling frame undercoverage, and nonresponse. A secondary objective is to reduce sampling errors, which is important because CHIS 2009 sample sizes within counties are fairly modest for some subclasses. In general, the sample is poststratified to as many independent figures as possible, subject to some constraints. In this discussion we use the term poststratification loosely and intend it to include raking, a form of multidimensional poststratification (see Brackstone and Rao, 1979). In CHIS 2009, the control totals are mainly derived from the 2009 California Department of Finance Population Estimates (State of California, Department of Finance, 2006a, 2006b), the 3-year 2008 American Community Survey (U.S. Census Bureau, 2010), and the Census 2000 Summary File 1 for California published by the U.S. Census Bureau (U.S. Census Bureau, 2001). Creation of the control totals at the person level is described in Chapter 7.

2.4 Combining Samples

In this section, we describe how the samples were combined to create the weights for CHIS 2009. Before explaining the approach for combining the samples, we examine the relationship between the different frames and samples.

Consider the different samples as illustrated in Figure 2-1, which shows the relationship for the Los Angeles County stratum.

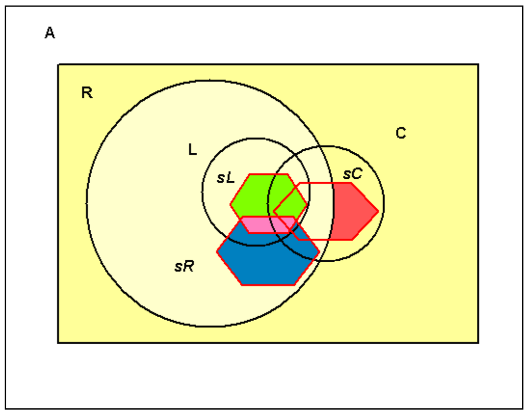


Figure 2-1. Landline, list, cell phone frames in CHIS 2009

Note: The figure is not drawn to scale. The sizes of the samples relative to the frames are smaller than shown in the figure.

Let A be all eligible households in Los Angeles County (represented by the large yellow rectangle in the diagrams). Let R (the large circle in the diagram) be all LA County households with telephone numbers in the landline frame, and L (smaller circle enclosed within R) be all LA County households with telephone numbers in the surname list frame. Note that by definition, R is included within A and that L is included within R (i.e., $L \subset R$). Let C be all LA County households with cell phone numbers, including those with no landline but with one or more cell phones (i.e., $C \cap \overline{R}$), and those with both types of telephone service (i.e., $C \cap R$). Notice that the cell frame, C, is not encompassed by R, but crosses both R and A. Let s_R, s_L, and s_C be households represented in the landline, surname list, and cell phone samples respectively.

Thus far the discussion has focused on households, but the sampling frames are actually of telephone numbers. Consider now the list sample s_L and the landline sample s_R. By definition, all numbers in the surname frame L are contained in the landline frame R, so all numbers on the surname frame have two probabilities of selection (one for the landline sample and the other for the surname sample). Since the landline/surname and cell phone frames themselves do not overlap (although the households they represent do, as shown in Figure 2-1), the inclusion of the cell phone sample does not affect the probability of selection of telephone numbers from the landline and surname samples. Thus, the landline and surname samples can be weighted following the same methods used in previous cycles of CHIS. That is, the base weights depend on whether or not the telephone number was found on the surname frame. (The information on about whether any of the landline sample cases were on the surname frames is available from the surname list vendor.) The expression of the base weights is described in more detail in Section 3.1.

As mentioned before, households with cell phone only, (i.e., $C \cap \overline{R}$) and households with both telephone services (i.e., $C \cap R$) were eligible in CHIS 2009. Their base weights were computed as the inverse of the probability of selection from the frame. Operationally, the cell-phone sample was weighted separately and at the same time as the landline sample applying the appropriate weighting adjustments.

Since the landline and cell phone populations and samples overlap and the drawn samples are probability samples, we used a multiple-frame estimation approach to combine and create weights for these samples. This approach followed the ideas of Hartley (1962) and was different from the approach used to combine the landline and surname samples. This method was needed because the multiple probabilities of selection of all units in the sample from both frames could not be determined.

There are three population domains of interest in the overlapping frames. The first domain called *a* includes all adults in households with only landline service, the second domain called *b* includes all adults in cell-only households, and the third domain called *ab* includes all adults in households with both landline and cell phones. Let Y be a characteristic for adults in a domain (e.g., the number of adults with health insurance). Let $\hat{Y}^A$ be the estimate of Y computed using the landline sample, and let $\hat{Y}^B$ the estimate of Y computed using the records in cell phone sample. An estimate of Y using the landline sample is

$$\hat{Y}^A = \hat{Y}_a^A + \hat{Y}_{ab}^A,$$

where $\hat{Y}_a^A$ is the estimate computed using the records from landline only households and $\hat{Y}_{ab}^A$ is the estimate computed using the adults with a landline and cell phone from in the landline sample. In a similar way, an estimate of Y based on the cell phone sample is $\hat{Y}^B = \hat{Y}_{ab}^B + \hat{Y}_b^B$ where $\hat{Y}_{ab}^B$ is the estimate computed using the adults with a landline and cell phone from the cell phone sample and $\hat{Y}_b^B$ is the estimate computed using the records from cell only households.

Notice that neither $\hat{Y}^A$ nor $\hat{Y}^B$ are unbiased estimates of Y. However, an unbiased estimate of Y can be computed as

$$\hat{Y} = \hat{Y}_a^A + \lambda \hat{Y}_{ab}^A + (1-\lambda)\hat{Y}_{ab}^B + \hat{Y}_b^B ,$$

where λ ($0 \le \lambda \le 1$) is the composite or weighting factor. In CHIS 2009, the value of λ was chosen to minimize the bias of $\hat{Y}$. The choice is outlined in Brick et al (submitted) and differs from the Harley approach that minimizes the variance. In either approach, the estimates $\hat{Y}_a^A$, $\hat{Y}_{ab}^A$, $\hat{Y}_{ab}^B$, and $\hat{Y}_b^B$ are poststratified before creating the composite estimator.

In CHIS 2009, a composite weight was created rather than the estimates. In this approach the value of λ is attached to the weights. The composite weights can be used to compute estimates for any variable (although the value of optimal value of lambda depends of the characteristic Y). For example, the expression for the estimate $\hat{Y}$ becomes

$$\hat{Y} = \sum_{i \in a \in A} w_i y_i + \sum_{i \in ab \in A} w_i \lambda y_i + \sum_{i \in ab \in B} w_i (1-\lambda) y_i + \sum_{i \in b \in B} w_i y_i .$$

Since the landline/surname and cell phone samples were independent samples, the estimates of variance can be computed using replication or linearization (i.e., Taylor series approximation).

In summary, the supplemental samples (i.e., geographic and surnames samples) were combined with the landline sample at the beginning of the weighting process. The cell phone sample and the combined landline-supplemental samples were first poststratified to telephone service control totals, combined through a composite factor, and then raked all together. Details of these adjustments are described in the following sections.

3. HOUSEHOLD WEIGHTING

For all CHIS samples, the first step in the weighting process for CHIS 2009 is creating a household weight for each completed screener interview. The household weight is not used for analytical purposes because the only data captured at the household level in the screener interview are for sampling purposes. However, the household weight is a key element for the computation of the person weights (i.e., adult, child, and adolescent).

This chapter is divided into eight sections, each describing the steps involved in creating the household weights. The first section reviews the creation of base weights. Subsequent sections describe the adjustments made to the base weights. These adjustments account for ported telephone numbers (numbers assigned to landline service that have been transferred to cell phones), sampled numbers that were not called, cases without full refusal conversion, unknown residential status, supplemental list sample eligibility, screener nonresponse, and households with multiple telephone numbers.

Knowledge of the sampling methods used in CHIS 2009 is essential to understanding the weighting procedures. We assume anyone interested in the weighting procedures is already familiar with the sampling approach – details are in *CHIS 2009 Methodology Series: Report 1 - Sample Design*.

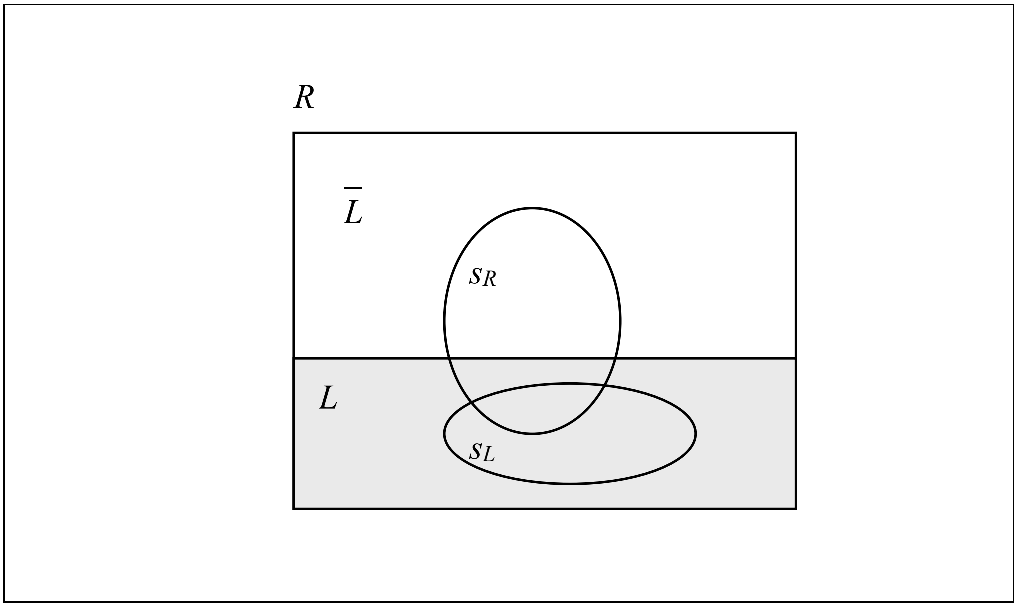
3.1 Base Weights

A base weight is created for each sampling unit in the different CHIS samples. For the landline, list, and cell samples, the sampling unit is the telephone number.

3.1.1 Landline and Surname List Base Weight

The base weight for the landline/list sample is computed as the inverse of the probability of selection of the telephone number. In CHIS 2009, telephone numbers were drawn from the landline frame and four mutually exclusive surname list frames (Korean, Vietnamese, Vietnamese and any other race but Korean, and Korean and any other race but Vietnamese). The base weights reflect the multiple probability of selection of telephone numbers between the landline and list frames.

Figure 3-1 shows the relationship between the landline frame and a single surname list frame for a single sampling stratum. The figure also shows the relationship between the landline and list samples drawn from each frame. In order to create the household base weights, we consider all landline telephone households in California as either being on the list (L) or as only being eligible for sampling from the landline sample ($\bar{L}$) as shown in Figure 3-1. The relationships are discussed in detail below.



* The figure is not drawn to scale. The sizes of the list frame (L) and list and landline samples (s_L and s_R) are smaller than shown in the figure.

Figure 3-1. Relationship between the landline frame (R), landline sample (s_R), list frame (L), and list sample (s_L) for a single stratum

The notation in the figure follows:

R	the landline frame containing all telephone numbers;
L	the list frame (i.e., surnames);
$\bar{L}$	all telephone numbers not found on the list – we assume that all the numbers in the list are found in R, and $R = L \cup \bar{L}$;
s_R	the simple random sample drawn from the frame R; and
s_L	the simple random sample drawn from the frame L.

We define the following:

N_R	the number of telephone numbers in the frame R;
N_L	the number of telephone numbers in the frame L;

n_R	the sample size (number of telephone numbers) of s_R; and
n_L	the sample size (number of telephone numbers) of s_L.

Notice that the landline sample s_R can be separated into two parts: s_{RL}, the portion of s_R that is found in the list (L), and $s_{R\bar{L}}$, the portion of s_R that is not found in the list $(\bar{L})$. The sample sizes for each portion are n_{RL} and $n_{R\bar{L}}$, respectively, and $n_R = n_{RL} + n_{R\bar{L}}$.

Consider L and $\bar{L}$ as two separate strata within the frame R. Since s_R is a simple random sample within R, the sample $s_{R\bar{L}}$ can be viewed as a simple random sample of size $n_{R\bar{L}}$ drawn from the $N_{\bar{L}}$ elements from stratum $\bar{L}$. Similarly, the sample s_{RL} can be viewed as a simple random sample of size n_{RL} drawn from the N_L elements from stratum L. In stratum L, there is a second sample s_L (the list sample). Since both samples s_L and s_{RL} are simple random samples, we can view them as a single sample of size $n_{RL} + n_L$ drawn from the N_L elements from stratum L. Notice that s_{RL} and s_L are not necessarily mutually exclusive; i.e., s_{RL} and s_L may contain duplicate telephone numbers. These numbers were removed from s_L during the sample selection.

The landline and list base weights can be expressed as follows:

- For sampled records that could only be sampled from the landline frame (landline numbers not found in the list L):

$$HHBW_{\bar{L}i} = \frac{N_{\bar{L}}}{n_{R\bar{L}}};$$

- For sampled records from the list and sampled records from the landline frame that are found in the list L (duplicate telephone numbers were eliminated from the list):

$$HHBW_{Li} = \frac{N_L}{n_{RL} + n_L}.$$

Creating these weights required classification of every telephone number by whether or not it was on the list irrespective of how it was sampled. It is easy to show that the resulting weights are composite weights derived by averaging the landline and list samples using a composite factor proportional to the sample sizes. Thus, this base weight produces an unbiased estimate in the traditional design-based framework.

The total telephone numbers in the landline frame and list frames (N_R and N_L) are computed separately. The landline sample was drawn using an RDD list-assisted approach from a stratified frame of 100 banks[3] with at least one listed telephone number in the state of California. Using this approach, a bank is drawn from the frame and two digits are randomly generated to complete the sampled telephone number. Therefore, the total number of telephone numbers in the landline frame in stratum *h*, N_{Rh}, is computed as

$$N_{Rh} = 100 \cdot NBANKS_h ,$$

where $NBANKS_h$ is the number of 1+ banks in March of 2009 in stratum *h*. A “1+” bank is defined as a 100 bank with at least one working telephone number.

Records on the list frames were assigned to landline sampling strata by linking telephone exchanges to the counties in the same way as for the landline sample. The list size by stratum $\left(N_{Lh}\right)$ is the number of records in the list assigned to stratum *h*.

As described in *CHIS 2009 Methodology Series: Report 1 - Sample Design,* the landline sample was drawn from strata defined as counties or groups of counties except for Los Angeles, San Diego, Orange, and Santa Clara. In Los Angeles County, 13 subsampling strata were created by the combination of areas with high/low concentration of Koreans and Vietnamese and eight Special Planning Areas (SPAs). Two substrata based on the concentration of Koreans and Vietnamese were created for San Diego, Orange, and Santa Clara Counties. The definition of the sampling strata and substrata, in addition to the number of telephone numbers in the landline frame, the number of sample cases, and base weights by frame type (landline, Korean only, Vietnamese only, Vietnamese and another group, and Korean, and another group lists), is shown in Appendix A, Tables A-1 and A-2. Table B-1 in Appendix B (rows 1.1 through 1.3) lists the sample counts, sums of base weights, and coefficients of variation by sampling stratum for these samples.

3.1.2 Cell Phone Base Weight

The cell phone sample was drawn for a stratified random sample of 1,000-series blocks dedicated to wireless service (NXXTYPE types 04, 55, 60) or PCS (personal

[3] A bank is defined as 100 consecutive telephone numbers with the same first eight digits including area code.

communication service types 65, 68)[4]. The cell sample base weight is similar to the landline base weight except that there are 1,000 numbers in each bank. The sampling stratum was defined by the area code of telephone numbers assigned to wireless service. For more details on the cell phone sample design, see *CHIS 2009 Methodology Series: Report 1 - Sample Design.*

Let $CPBW_{hi}$ be the base weight for the *i*-th sampled cell phone number in the *h*-th stratum (defined by area codes), the base weight is computed as

$$CPBW_{hi} = \frac{NC_h}{n_h},$$

where n_h is the total sampled numbers in stratum h, and NC_h is the total numbers in stratum h, computed as $NC_h = 1000 \cdot NS_h$ where NS_h is the number of 1,000 blocks in stratum *h*. Note that the stratum definition for the cell phone sample is different from that of the landline sample. The definition of sampling strata, the number of telephone numbers in the frame, the number of sampled cases, and average base weights are shown in Appendix A, Table A-2.

3.2 Ported Telephone Number Adjustment

Telephone numbers sampled as part of the landline sample but identified as cell phone numbers during the purging process were dialed as part of the cell sample in CHIS 2009. However, since the cell sample target was met before the end of data collection, some ported telephone numbers were not dialed. The weights of ported numbers that were dialed were adjusted to account for ported numbers from the landline sample that were not dialed. In this adjustment, the dialed ported numbers were assumed to be a random sampled of all ported telephone numbers identified in the landline sample. The ported telephone adjusted weight, $HHA1W_i$, is computed as

$$HHA1W_i = HHA1F_i * HHBW_i,$$

where $HHA1F_i$ is the ported telephone number adjustment factor computed as:

[4] There are some additional technical restrictions in the sampling, such as making sure the number can be dialed into and that toll-free numbers are excluded.

$$HHA1F_i = \begin{cases} \dfrac{\sum_{i \in CDIALED, CN\_DIALED} HHBSW_i}{\sum_{i \in CDIALED} HHBSW_i} & \text{If } i \in CDIALED \\ 0 & \text{If } i \in N\_CDIALED \end{cases},$$

where the group *CDIALED* denotes dialed cell phone numbers identified as ported in the landline sample , *N_CDIALED* denotes those that were not dialed. This adjustment is done separately by sampling stratum. This adjustment was not applied to the cell phone sample and the adjustment factor $HHA1F_i$ was set to one for all records in this sample. Table B-1 in Appendix B (rows 2.1 through 2.4) shows the sum of weights before and after the adjustment

3.3 New Work Adjustment

Telephone numbers were drawn during data collection depending on the number of completed interviews achieved so far and the projected number of completed interviews at the end of the data collection period. However, not all new drawn telephone numbers were dialed because the targets of some strata were met before exhausting the sample. In this adjustment, the weights were adjusted to account for those cases that were not dialed. The dialed telephone numbers were assumed to be a random sample of all drawn telephone numbers. The new work telephone adjusted weight, $HHA2W_i$, is computed as

$$HHA2W_i = HHA2F_i * HHA1W_i \,,$$

where $HHA2F_i$ is the new work adjustment factor computed as:

$$HHA2F_i = \begin{cases} \dfrac{\sum_{i \in DIALED.N\_DIALED} HHA1W_i}{\sum_{i \in DIALED} HHA1W_i} & \text{If } i \in DIALED \\ 0 & \text{If } i \in N\_DIALED \end{cases},$$

where the group *DIALED* denotes dialed telephone numbers and *N_DIALED* denotes those that were not. This adjustment is very small and was done separately by sampling stratum and mailable status. This adjustment was applied to telephone numbers in the landline and list samples. The adjustment factor $HHA2F_i$ was set to one for all records in the cell phone sample. Table B-1 in Appendix B (rows 3.1 through 3.4) shows the sum of weights before and after the adjustment.

3.4 Refusal Conversion Adjustment

Subsampling for refusal conversion was a technique used in CHIS 2003, 2005, and 2007. It was used to gain cooperation in households in which a member refuses to participate in the study at the screener level by shifting some resources from the less productive, labor-intensive task of refusal conversion to the more productive task of completing extended interviews. Due to changes over time in the relative efficiency of different kinds of work, subsampling for refusal conversion was not implemented in CHIS 2009 and all refusals in the landline and surname samples were eligible for two refusal conversion attempts. However, towards the end of the field period, additional telephone numbers were released in selected strata to meet the targets for the number of completed interviews. In some instances, no or only one refusal conversion was attempted because the protocol could not be implemented before the end of the data collection period. In this adjustment, the weights of the cases with two refusal conversion attempts were adjusted to account for those cases that have none or only one refusal conversion. It is assumed that refusals without refusal conversion attempts were a random sample of those with refusal conversion attempts in this adjustment.

This adjustment did not affect cases from the surname samples because all the refusals in these samples had two conversion attempts in the standard protocol. This adjustment was not applicable to households in the cell phone sample because only one refusal conversion was attempted, and this protocol was implemented for all numbers that were dialed. Therefore this adjustment factor was set to one for those samples.

Before adjusting the weights for screener interview refusal subsampling, telephone numbers were classified into screener refusal groups using their refusal status (i.e., whether the respondent ever refused) and the value of the refusal conversion flag as shown in Table 3-1.

Table 3-1. Screener refusal groups for landline sample

Screener refusal group	Respondent ever refused screener interview?	First Refusal Subsampling Flag	Second Refusal Subsampling Flag	Description
NRef	No	N/A	N/A	Households where respondent did not refuse the screener interview (includes complete and incomplete screener interviews)
RefC1	Yes	Yes	No	Households where respondent refused the screener interview and only first refusal conversion procedures were used
RefC2	Yes	Yes	Yes	Households where respondent refused the screener interview and both first and second refusal conversion procedures were used
RefNC	Yes	No	No	Households where respondent refused the screener interview and refusal conversion procedures were not used

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

The first refusal subsampling adjusted weight, $HHA3W_i$, is:

$$HHA3W_i = HHA3F_i * HHA2W_i$$

where $HHA3F_i$ is the first refusal subsampling adjustment factor computed as:

$$HHA3F_i = \begin{cases} \dfrac{\sum\limits_{i \in (RefC1, RefNC)} HHA2W_i}{\sum\limits_{i \in RefC1} HHA2W_i} & \text{If } i \in RefC1 \\ 0 & \text{If } i \in RefNC \\ 1 & \text{If } i \in RefNC \end{cases},$$

where the groups $RefC1$, $RefNC$, and $NRef$ are defined in Table 3-1, $HHA2W_i$ is the new work adjusted weight, and $\delta_i(c)$ is 1 if the number is in sampling stratum c and is zero otherwise.

The second refusal subsampling adjusted weight, $HHA4W_i$, is:

$$HHA3W_i = HHA4F_i * HHA3W_i$$

where $HHA2F_i$ is the second refusal subsampling adjustment factor computed as:

$$HHA4F_i = \begin{cases} \dfrac{\sum_{i \in (RefC2, RefNC1)} HHA3W_i}{\sum_{i \in RefC2} HHA3W_i} & \text{If } i \in RefC2 \\ 0 & \text{If } i \in RefNC \\ 1 & \text{If } i \in NRef \end{cases},$$

where the groups $RefC2$, $RefNC$, and $NRef$ are defined in Table 3-1. Table B-1 in Appendix B (rows 4.1 through 5.4) shows the sum of the weights before and after the refusal conversion subsampling adjustments.

3.5 Unknown Residential Status Adjustment

Telephone numbers with unknown residential status are those that could not be classified by residential status at the end of data collection despite being dialed many times. They are telephone numbers with only answering machine results or some combination of answering machine and ring no answer results (screener disposition code of *NM*) or all ring no answer results (screener disposition of *NA*). Before adjusting the weights to account for telephone numbers with unknown residential status, the proportion of eligible residential telephone numbers among those numbers with unknown residential status was estimated. This estimate was also used in the computation of the response rates described in *CHIS 2009 Methodology Series: Report 4 - Response Rates*.

In CHIS 2009, the estimated proportion of unknown residential telephone numbers considered residential (p_{res}) was computed separately for the landline, surname, and cell phone samples. The proportion p_{res} was computed following the CASRO (1982) recommendation as the proportion of the resolved or observed sample units that are residential. Since telephone numbers were sampled with different selection probabilities and were adjusted differentially for refusal conversion, the weighted number of telephone numbers were used rather than the number of cases (unweighted) to compute p_{res}. This use of weights also compensates for the under- and oversampling implemented in different geographic areas.

Table 3-2 shows the values of p_{res} for the landline sample, calculated separately for each combination of mail status, urbanicity, and how the answering machine result was coded by interviewers. As expected, the estimated proportion of residential households is much lower for answering machines coded as "possible nonresidential" compared to those coded as "possible residential." For example, in urban strata among mailable cases, the estimated proportion of residential households with mailable addresses and answering machine results coded as possible residential is 94.0 percent, while the estimated proportion of those coded possible nonresidential is 20.4 percent. The lowest percentages of residential telephone numbers are for the numbers that were not mailable and had answering machine messages coded as possible nonresidential or unknown.

Table 3-2. Estimated residential proportion for the landline sample

Mail status	Urban status	Answering machine code	p_{res}
Mailable	Urban	No machine	0.673
Mailable	Urban	Possible residential	0.940
Mailable	Urban	Possible nonresidential	0.204
Mailable	Urban	Unknown	0.877
Mailable	Not urban	No machine	0.749
Mailable	Not urban	Possible residential	0.944
Mailable	Not urban	Possible nonresidential	0.268
Mailable	Not urban	Unknown	0.901
Not mailable	Urban	No machine	0.260
Not mailable	Urban	Possible residential	0.887
Not mailable	Urban	Possible nonresidential	0.101
Not mailable	Urban	Unknown	0.018
Not mailable	Not urban	No machine	0.293
Not mailable	Not urban	Possible residential	0.902
Not mailable	Not urban	Possible nonresidential	0.131
Not mailable	Not urban	Unknown	0.020

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

Table 3-3 shows the values of p_{res} for the list samples. Since there were no differences by type of list sample, the values of p_{res} were computed combining the cases from the lists.

Table 3-3. Estimated residential proportion for the list samples

Answering machine code	p_{res}
No machine	0.507
Answering machine possible residential	0.927
Answering machine possible nonresidential	0.149
Answering machine unknown	0.073

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

Table 3-4 shows the values of p_{res} for the cell phone sample. This proportion was computed by type of phone (ported or wireless assigned) and region. There were differences in the estimated proportions by these groups; the value of p_{res} for ported cell phones is on average 18 percentage points higher than the p_{res} for wireless assigned cell phones.

Table 3-4. Estimated residential proportion for the cell phone samples by type of cell phone and region

Type of cell phone	Region	p_{res}
Ported	1 - Northern & Sierra Counties	0.623
	2 - Greater Bay Area	0.626
	3 - Sacramento Area	0.615
	4 - San Joaquin Valley	0.564
	5 - Central Coast	0.760
	6 - Los Angeles	0.386
	7 - Other Southern California	0.711
Wireless assigned	1 - Northern & Sierra Counties	0.517
	2 - Greater Bay Area	0.640
	3 - Sacramento Area	0.667
	4 - San Joaquin Valley	0.586
	5 - Central Coast	0.575
	6 - Los Angeles	0.613
	7 - Other Southern California	0.548

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

The estimated proportion of residential households among the unknown residential telephone numbers or addresses p_{res} is then used to adjust the weights for unknown residential status. The residential status adjusted weight, $HHA5W_i$, is

$$HHA5W_i = HHA5F_i * HHA4W_i,$$

where $HHA5F_i$ is the unknown residential status adjustment factor computed as:

$$HHA5F_i = \begin{cases} \dfrac{\sum_{i \in RES} HHA4W_i + \sum_{i \in UNK\_RES} p_{res} * HHA4W_i}{\sum_{i \in RES} HHA4W_i} & \text{If } i \in RES \\ 0 & \text{If } i \in UNK\_RES \end{cases},$$

where the group *RES* denotes telephone numbers identified as residential and *UNK_RES* denotes telephone numbers with unknown residential status.

This adjustment is done separately by sample type. In the landline sample, the adjustment is done within sampling stratum by mailable status. In the list sample, the adjustment is by surname list type. This adjustment was not applied to the cell phone sample even though Table 3-4 shows that there were differences between the ported and wireless assigned cell phones. The adjustment factor $HHA3F_i$ was set to one for all records in this sample. Table B-1 in Appendix B (rows 6.1 through 6.5) shows the sum of weights before and after making the adjustment for unknown residential status for the landline, surname, and cell sample.

3.6 Sample Eligibility Nonresponse Adjustment

After the unknown residential status adjustment, the weights are adjusted for eligibility in the samples where screening is used to identify eligible respondents. In CHIS 2009, screening was used only to identify households with adults of Korean or Vietnamese descent in the surname samples. Therefore, this adjustment is only applicable to these samples.

The weights were adjusted to account for households in which the ethnic origin of the adults (i.e., whether Korean or Vietnamese) could not be determined. Telephone numbers from the surname list samples were eligible only if one or more adults in the household considered themselves of Vietnamese or Korean descent[5].

Households with at least one adult from one of these ethnic groups are referred to as "list-eligible" households. If a household from the supplemental sample was found to be list-eligible, then one adult from these ethnic groups was selected for the extended interview. If the household was not list-eligible (i.e., no adults of Vietnamese or Korean descent), then the screener interview was terminated and the case was coded as a list-ineligible.

Screening on eligibility and retaining only list-eligible households in the surname supplemental samples was a relatively efficient method for increasing the number of Korean and Vietnamese extended interviews in CHIS 2009 and previous cycles. The information on the ethnic origin of the adults was used to avoid unnecessary interviews of adults with a different ethnic origin, who were represented adequately in the landline sample.

[5] Question SC6A1 of the screener interview asked, "Do any of these adults who live in your household consider themselves to be Korean or Vietnamese or of Korean or Vietnamese descent?"

Household list eligibility could not be determined for nonresponding households in the surname list samples, and the weights had to be adjusted for unknown list eligibility. The weights of the households with unknown list eligibility were distributed between the list-eligible and ineligible households in the surname samples. The assumption in this adjustment was that the proportion of list-eligible/ineligible households among the households with unknown list eligibility was the same as the observed proportion in the sample with known eligibility. The cases were classified in response groups as indicated in Table 3-5.

Table 3-5. List eligibility response groups

List eligibility response status group		Description
L_E	List-eligible	Household from the surname list sample with at least one list-eligible adult (i.e., adult of Korean or Vietnamese descent).
L_IN	List-ineligible	Household from the surname list sample without any list-eligible adult (i.e., no adults of Korean or Vietnamese descent).
L_UNK	List eligibility unknown	Household from the surname list sample where the eligibility of the adults could not be determined.
L_NA	List eligibility not screened	Household from all other samples (not screened for eligible ethnicity).

The list eligibility nonresponse adjusted household weight, $HHA5W_i$, is computed as

$$HHA6W_i = HHA6F_c * HHA5W_i ,$$

where $HHA6F_c$ is the list eligibility nonresponse adjustment factor computed as

$$HHA6F_c = \begin{cases} \dfrac{\sum_{i \in L\_E, L\_IN, L\_UNK} HHA5W_i \delta(c)}{\sum_{i \in L\_E, L\_IN} HHA5W_i \delta(c)} & \text{If } i \in L\_E, L\_IN \\ 0 & \text{If } i \in L\_UNK \\ 1 & \text{If } i \in L\_NA \end{cases}$$

,

where the groups *L_E*, *L_IN*, *L_UNK*, and *L_NA* are defined in Table 3-4, and $\delta_i(c)$ is 1 if the number is in list eligibility nonresponse adjustment cell *c* and is zero otherwise. The nonresponse adjustment cells correspond to the list sample type (i.e., Korean, Vietnamese, Korean-other, and Vietnamese-other samples). Table B-1 in Appendix B (rows 7.1 through 7.4) shows the sum of weights before and after the list eligibility nonresponse adjustment.

3.7 Screener Nonresponse Adjustment

In this step, the household weight is adjusted to account for households that did not complete the screener interview. The nonresponse adjustment cells were created separately for the main landline and surname list samples and utilized information on the presence of children in the household from the screener[6].

In the first step of screener nonresponse adjustment we adjusted the weights to account for the presence of children in the household. The weights of nonresponding households with a known child status were distributed to responding households. This weight, $HHA7W_i$, is:

$$HHA7W_i = HHA7F_c * HHA6W_i,$$

where $HHA7F_c$ is the unknown presence of children adjustment factor computed as

$$HHA7F_i = \begin{cases} \dfrac{\sum\limits_{i \in SC\_KCS, SC\_UCS} HHA6W_i \delta_i(c)}{\sum\limits_{i \in SC\_KCS} HHA6W_i \delta_i(c)} & \text{If } i \in SC\_KCS \\ 0 & \text{If } i \in SC\_UCS \end{cases}$$

where the group SC_KCS is the set of screener respondents with known child presence status, and SC_UCS is the set of screener nonrespondents with unknown child status, and $\delta_i(c)$ is 1 if the number is in screener nonresponse adjustment cell c and is zero otherwise. Table B-1 in Appendix B (rows 8.1 through 8.4) shows the sum of weights before and after the unknown presence of children in household adjustment.

In the second step of screener nonresponse adjustment we adjusted the weights to account for screener nonresponse among households with a known presence of children. This weight, $HHA8W_i$, is:

$$HHA8W_i = HHA8F_c * HHA7W_i,$$

[6] There are differences in response rates between household with and without children. See *CHIS 2009 Methodology Series: Report 4 – Response Rates*

where $HHA8F_c$ is the screener nonresponse adjustment factor computed as

$$HHA8F_c = \begin{cases} \dfrac{\sum\limits_{i \in SC\_R, SC\_NR} HHA7W_i \delta_i(c)}{\sum\limits_{i \in SC\_R} HHA7W_i \delta_i(c)} & \text{If } i \in SC\_R \\ 0 & \text{If } i \in SC\_NR \end{cases},$$

where the group *SC_R* is the set of screener respondents, and *SC_NR* is the set of screener nonrespondents, and $\delta_i(c)$ is 1 if the number is in screener nonresponse adjustment cell c and is zero otherwise.

List-ineligible households (i.e., households with no adults of Korean or Vietnamese origin) from the surname list samples (group *R_IN* defined in the previous section) were considered as screener nonrespondents (group *SC_NR*) in this adjustment. Although these cases were households with only list-ineligible adults, they still represented households with eligible adults for the landline sample extended interview who were screened out. Table B-1 in Appendix B (rows 9.1 through 9.4) gives the sum of weights before and after the screener nonresponse adjustment.

3.8 Multiple Telephone and Duplicate Respondent Adjustments

At the end of the screener interview for the landline sample, information about the existence of additional telephone numbers and their use in the household was collected. If more than one telephone number is used for residential purposes (not solely for business, fax or computer use, etc.), the household has a greater probability of selection because it could have been selected through any of the additional telephone numbers in the household. In such cases, the household weight is adjusted to reflect the increased probability of selection. The multiple telephone adjusted household weight, $HHA9W_i$, is computed as:

$$HHA9W_i = HHA9F_c * HHA8W_i,$$

where $HHA9F_c$ is the multiple telephone adjustment factor computed as:

$$HHA9F_c = \begin{cases} 0.5 & \text{If household } i \text{ has more than one residential telephone number} \\ 1 & \text{Otherwise} \end{cases}.$$

In this adjustment, we assume that there is at most one additional residential-use telephone number in the household. Table B-1 in Appendix B (rows 10.1 through 10.5) identifies the sum of weights before and after the multiple telephone adjustment. This adjustment was not applied to the cell sample and therefore the adjustment factor was set to 1 in this step.

After adjusting the weights for the increase probability of selection due to multiple telephones, the weights were first adjusted for households that were sampled through different landline telephones (i.e., different telephone numbers for the same household). Since respondents were not interviewed twice, the second attempted interview was coded as a duplicate respondent. Since these numbers represent the same household, the weight of the first interview is adjusted to account for the second attempted interview. In this step, the weight for the duplicate was distributed to the completed screener. The duplicate respondent adjustment factor $ODF1_i$ was computed as:

$$ODF1_i = \begin{cases} \dfrac{HHA9W_{Complete} + HHA9W_{Duplicate}}{HHA9W_{Complete}} & \text{Landline completed interview with duplicate} \\ 0 & \text{Lnadline duplicate respondent} \\ 1 & \text{Otherwise} \end{cases}$$

In other cases, respondents were contacted by different telephone types (landline and cell phone). In these cases the weight of the duplicate respondent was distributed to the non-duplicate numbers within sampling strata. In this case, the second duplicate respondent adjustment factor $ODF2_i$ was computed as:

$$ODF2_1 = \begin{cases} \dfrac{\sum_{i \in c} HHA9W_i}{\sum_{i \in c \text{ and it is not duplictae}} HHA9W_i} & \text{Duplicate respondent with completed interveiw} \\ 0 & \text{Duplicate respondent} \\ 1 & \text{Otherwise} \end{cases} .$$

The household weight adjusted for duplicate respondents, $HHA10W_i$, is computed as

$$HHA10W_i = HHA10F0_i * HHA9W_i .$$

where the overall duplicate respondent factor $HHA10F_i$ adjustment factor was computed as $HHA10F0_i = ODF1_i * ODF2_i$. Table B-1 in Appendix B (rows 11.1 through 11.2) identifies the sum of weights before and after this adjustment.

4. ADULT WEIGHTING

An adult final weight was created for each adult who completed the adult extended interview.[7] The initial adult weight is the product of the final household weight and the reciprocal of the probability of selecting the adult from all adults in the household for the landline (including the surname list) sample. For the cell phone sample, the initial weight is the product of the final household weight and the number of adults in the household where the cell phone is shared; if the cell phone is not shared, the initial adult weight equals the final household weight. In subsequent steps, the initial adult weight is adjusted for nonresponse. Before raking the weights to known control totals, the landline and cell phone samples are poststratified to controls by telephone type. After this step, a composite weight that combines the landline and cell phone sample was created. Undercoverage of adults that could not be interviewed because they reside in households without a landline or cell phone was compensated for by a raking adjustment that included a dimension to reduce the undercoverage bias. Details on creating the adult weights follow.

4.1 Adult Initial Weight

As described in *CHIS 2009 Methodology Series: Report 1 - Sample Design*, one adult was sampled with equal probability from all adults in the household using the Rizzo method (see Rizzo et. al., 2004). The initial adult weight is the product of the final household weight and the inverse of the probability of selection of the adult. The expression for the adult initial weight, $ADA0W_j$, is

$$ADA0W_i = ADCNT_i \cdot HHA9W_i \,,$$

where $ADCNT_i$ is the total number of adults in household *i* for the landline and surname samples and the number of adults in the household (if there are adults that share the cell phone) for the cell phone sample, and $HHA9W_i$ is the multiple telephone adjusted weight described in the previous chapter.

This scheme for the cell phone sample assumes that, in cell phone households with more than one adult, each adult has a cell phone (or shares a different cell phone) if the sampled cell phone is not shared. If the cell phone is shared, it assumes that all adults in the household share

---

[7] Adult extended interviews are considered complete provided the adult completed through Section K on employment and income.

the same phone. Appendix B, Table B-2 (rows 1.1 through 1.3) shows the number of adults, sum of initial weights, and coefficient of variation for the landline and cell samples for the state.

4.2 Adult Nonresponse Adjustment

Regardless of the sample (i.e., landline or cell phone samples), some households completed the screener interview but the sampled adult did not complete the extended adult interview. In addition, in a few cases it was discovered during the extended interview that the sampled person was under 18 years of age and hence ineligible. To account for both sampled adults who did not complete the extended interview and for ineligible sampled persons, the adult initial weight was adjusted for extended interview nonresponse. Prior to making the adjustment, extended interviews were classified into response groups as indicated in Table 4-1.

Table 4-1. Extended interview response groups

Response status group		Description
ER	Eligible respondent	Adult who completed the extended interview
IN	Ineligible	Ineligible person
UNK	Unknown eligibility	Sampled adult could not be contacted and eligibility verified for extended interview

The adult nonresponse adjusted weight, $ADA1W_i$, is

$$ADA1W_i = ADA1F_c \cdot ADA0W_i ,$$

where $ADA1F_c$ is the adult nonresponse adjustment factor given by

$$ADA1F_c = \begin{cases} \dfrac{\sum_{i\in(ER,IN,UNK)} ADA0W_i \cdot \delta_i(c)}{\sum_{i\in(ER,IN)} ADA0W_i \cdot \delta_i(c)} & i \in (ER, IN) \\ 0 & i \in UNK \end{cases},$$

where *ER, ENR* and *IN* are defined in Table 4-1, *c* indicates the adult extended interview nonresponse adjustment cell, and $\delta_i(c) = 1$ if the adult belongs to cell *c* and is zero otherwise.

Table 4-2 lists the variables that were considered in defining the nonresponse adjustment cells. All of these have been examined in previous CHIS cycles. A nonresponse analysis showed that sex, child-first interview status, age group, and whether the sampled adult was also the screener respondent were the best variables for creating nonresponse cells. Nonresponse cells with fewer than 30 respondents or with large adjustment factors were combined with adjacent cells. All the cells were created within sampling stratum. Appendix B, Table B-2 (rows 2.1 through 2.5) shows the sum of weights before and after the nonresponse adjustment for the landline and cell phone samples. Ineligible persons were dropped following this weighting step.

Table 4-2. Variables used for the creation of nonresponse adjustment cells for the adult weights

Variable	Levels	
Sex of adult respondent	1.	Male
	2.	Female
Child-first interview	1.	Child-first procedures in affect
	2.	Child-first procedures not in affect
Adult age group	1.	18-30 years old
	2.	31-45 years old
	3.	46-65 years old
	4.	65 years or older
Adult screener Respondent	1.	Sampled adult was screener respondent
	2.	Sampled adult was not screener respondent

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

4.3 Composite Weight

The next step in weighting was to combine the landline and cell samples. Before creating the composite weights, both samples were poststratified separately to control totals defined by telephone service or type (i.e., persons in landline only households, persons in cell phone only households, and persons in households with both services). The distribution of telephone usage for California was derived from the National Health Interview Survey for January to June 2010 for the West region. The poststratified person weight, $PPERW_j$ is computed as

$$PPERW_j = \frac{TEL\_USAGE\_CT_i}{\sum PERW_j} * PERW_j$$

where $PERW_j$ is the person weight (i.e., adult, child, or adolescent) and $TEL\_SERVICE\_CT_i$ is the control total by telephone service (landline only, cell phone only, both services). Appendix B, Table B-5 (rows 1.1 through 1.3) shows the sum of weights before and after this adjustment.

Once the samples were poststratified, a composite weight that combined the landline and cell phone sample was created. Based on research by Brick et. al (2010), the composite factor

$\lambda = 0.9$ was used to reduce the bias of estimates computed from the combined sample. This factor and its complement ($1-\lambda$) can be seen as additional weighting adjustment factors to apply to the poststratified weights. The expression of the composite weight, $COMBW_j$, is

$$COMBW_j = \begin{cases} PPERW_j & \text{If person i lives in a household with cell only or landline only} \\ \lambda * PPERW_j & \text{If person i lives in a household with cell and landline from the landline sample} \\ (1-\lambda) * PPERW_j & \text{If person i lives in a household with cell and landline from the cell sample} \end{cases}$$

where $PPERW_j$ is the poststratified person weight above. Appendix B, Table B-5 (row 2.1) shows the sum of weights before and after this adjustment.

4.4 Adult Trimming Factors

Before benchmarking the adult weights to the known total of adults in California in 2009, we examined the distribution of the composite weights to determine if there were very large weights that could have a large effect on either the estimates or the variances of the estimates. When observations with large weights were found, the weights for these cases were reduced in a process called trimming.

As in previous cycles, we computed statistics to identify influential weights that were candidates for trimming. These statistics and other variations were studied in detail in Liu et al. (2004). The first statistic is a function of spacing of the weights. Let $w_{(1)},\ldots,w_{(n)}$ be the order statistics for the adult weights $w_1,\ldots,w_n$ and define "spacing" z_i as the distance (difference) between a ranked weight $w_{(i)}$ and the next ranked weight $w_{(i-1)}$ (i.e., $z_i = w_{(i)} - w_{(i-1)}$). The statistic $d5\_space_i$ for a ranked $w_{(i)}$ is defined as

$$d5\_space_i = \frac{z_i}{z_{i-1} + z_{i-2} + z_{i-3} + z_{i-4} + z_{i-5}}.$$

The second statistic used computes the distance between a weight and the next largest weight relative to the size of the weight. The statistic is

$$rel\_space_i = \frac{z_i}{w_{(i)}} \times 10.$$

We also computed a third statistic defined as

$$diff\_dist_i = distance_i - distance_{i-1},$$

where $distance_i$ is the relative distance for the weight $w_{(i)}$ computed as

$$distance_i = \frac{\left|w_{(i)} - median(\mathbf{w})\right|}{MAD},$$

where $\mathbf{w} = (w_i,...,w_n)^t$ and the median absolute deviation $MAD = median(\left|w_i - median(\mathbf{w})\right|)$.

The three statistics for the largest 20 weights in each stratum were examined separately. When all three statistics were greater than 1 then the case was a primary candidate for trimming. The final decision on trimming involved the inspection of the weight distribution within sampling stratum.

The trimmed weight $TRMW_i$ is computed as

$$TRMW_i = TFACT_i * PPERW_i,$$

where $TFACT_i$ is the trimming factor for the sampled adult *i* given by

$$TFACT_i = \begin{cases} 1 & \text{if the weight } i \text{ is not trimmed} \\ t_i & \text{otherwise} \end{cases}$$

where $0 < t_i < 1$.

For the adult extended interview 70 records were trimmed[8]. The trimming factor ranged between 0.0790 and 0.9996. Table B-5 (rows 2.1 and 3-1 to 3-3) shows the strata with trimmed weights by self-reported and the sum of weights before and after trimming for the different weights.

[8] The trimming was done prior to the raking adjustment; however, it was an iterative process. After the trimming and raking, the distribution of the weights was re-examined, and new decisions were made about trimming. This might have changed the decision about which weights should be trimmed or the magnitude of the trimming factor. If the decision was made, the trimmed and raked weights were discarded and new trimming and raking were undertaken. The number of trimmed weights reported here is at the completion of the overall process.

4.5 Adult Raked Weight

The next step in the adult weighting was raking the trimmed weights to population control totals to produce estimates consistent with the 2009 California Department of Finance (DOF) Population Estimates. Included in the raking adjustment is an undercoverage adjustment for adults in households without a - telephone discussed earlier. The specific control totals and the method used to create them are described in Chapter 7.

Raking is a commonly used estimation procedure in which estimates are controlled to marginal population totals. It can be thought of as a multidimensional poststratification procedure because the weights are poststratified to one set (a dimension) of control totals, then these adjusted weights are poststratified to another dimension. The procedure continues until all dimensions are adjusted. The process is then iterated until the control totals for all dimensions are simultaneously satisfied (at least within a specified tolerance). Raking is also described in more detail in Chapter 7.

The adult raked weight, $RAKEDW_i$, can be expressed as

$$RAKEDW_i = TRMW_i \cdot \prod_{k=1}^{K} RAKEDF_{k_l} \ ,$$

where $RAKEDF_{k_l}$ is the raking factor for dimension *k*, level *l* in which adult *i* is. For example, if the 4th dimension ($k=4$) is sex with two levels (l=1 for male and l=2 for female), then the raking factor for this dimension is $RAKEDF_{4_1}$ for the adult male. The raking factors are derived so the following relationship holds for every raking dimension *k*, and level *l*,

$$CNT_{k_l} = \sum_i \delta(k_l)_i \cdot RAKEDW_i \ ,$$

where CNT_{k_l} is the control total, and $\delta(k_l)_i = 1$ if the adult *i* is in level *l* of dimension *k* and zero otherwise.

The weights, which include landline and cell sample weights, were raked to known control totals for California. Table B-5 (rows 1.3 and 2.2) shows the sum of weights before and after this raking adjustment.

5. CHILD WEIGHTING

A final child weight was created for all completed child extended interviews. In CHIS 2009, unlike CHIS 2007, children (and adolescents) were selected in the cell phone sample. The steps for the child weighting are similar to those for adults described in the previous chapter. One exception is an additional weighting adjustment needed to account for nonresponse in a section of the adult interview where the majority of the children were sampled. A more complete discussion of this adjustment is found in Section 5.1. The format of this chapter follows that for the adult weighting, with the creation of the child initial weights and the adjustments for nonresponse, telephone service poststratification, composite weight, trimming, and finally raking.

5.1 Household-Level Adjustment

The main difference between the child (and adolescent) weighting procedures and those for adults is that adults were always sampled in the screener. Children and adolescents could be selected at the end the screener interview or in Section G of the adult extended interview. The selection of children at the end of the screener interview is called the child-first procedure. Weights for children and adolescents selected in Section G must be further adjusted to account for nonresponse at the adult interview level. On the other hand, weights of the child-first children and adolescents were not adjusted for adult nonresponse.

Telephone numbers were classified into completion groups (*SECGST*) by Section G completion status and their child-first interview status as shown in Table 5-1.

Table 5-1. Section G completion groups

Section G completion group (*SECGST*)	Child-first interview?	Section G completed by adult?	Description
C1st	Yes	N/A	Households where the child-first interview procedures occurred
NC1stGC	No	Yes	Households where the child-first interview procedures did not occur and section G was completed
NC1stGNC	No	No	Households where the child-first interview procedures did not occur and section G was not completed

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

To account for adults who did not complete Section G of the adult interview (hence, no child or adolescent could be sampled), the household final weight $HHA9W_i$ was adjusted. We refer to this adjusted weight as the Section G adjusted household weight, $HHA10W_i$, and it is

$$HHA10W_i = HHA10F_c * HHA9W_i ,$$

where

$$HHA10F_i = \begin{cases} \dfrac{\sum_{i \in NCstGC, NC1stGNC} HHA9W_i \delta(c)}{\sum_{i \in NCstGC} HHA9W_i \delta(c)} & \text{If } i \in NCstGC \\ 0 & \text{If } i \in NC1stGNC \\ 1 & \text{If } i \in C1st \end{cases} ,$$

and where the section G completion groups *C1st*, *NC1stGC*, and *NC1stGNC* are defined in Table 5-1, c denotes the Section G nonresponse adjustment cell, and $\delta_i(c) = 1$ if the adult belongs to cell c and is zero otherwise. Following this adjustment, the weights were positive for all households with sampled adults who completed section G and either completed, partially completed, or did not complete the adult interview[9]. Note that this adjustment can be considered as a household adjustment in addition to the household weight. Table B-1 in Appendix B (rows 12.1 through 12.2) identifies the sum of weights before and after this adjustment.

The Section G nonresponse adjustment cells were created within sampling strata using a combination of the mailable status (known address/mailed letter, unknown address) and the presence of children and/or adolescents. The information about the presence of children and adolescents is collected during the screener interview.

5.2 Initial Child Weight

The initial child weight is the product of the adjusted household weight and the probability of sampling the child within the household. The selection of the child was done in two steps. In the first step, one adult was randomly selected among all adults in the household. In the second step, one child was randomly selected among all the children associated with the sampled adult (i.e., the sampled adult is the parent or legal guardian of the child). If the sampled adult did

[9] If the adult interview was not completed or was only partially completed in a case that did not use the child-first procedure, no attempt was made to complete a child/teen interview.

not have an associated child, then no child was sampled even if there were children present in the household. See *CHIS 2009 Methodology Series: Report 1 - Sample Design* for information on the within-household person selection process.

Since the selection of a child within a household depends on the relationships among children and adults within the household, these relationships were defined before sampling children. The probability of selection reflects the fact that the sampled child could have been selected through the spouse/partner of the sampled adult if both are the parents or legal guardians[10] of the sampled child. Accordingly, the initial child weight, $CHA0W_j$, is

$$CHA0W_j = \frac{1}{CHPROB_j} HHA10W_i ,$$

where $HHA10W_i$ is the section G adjusted weight, $CHPROB_j$ is the probability of selecting the j^{th} child associated with the i^{th} sampled adult and is relatively complex. If the sampled adult does not have a spouse/partner living in the household or if the spouse/partner of the sampled adult is not the parent or legal guardian of the sampled child, then

$$CHPROB_j = \frac{1}{ADLTCNT} \cdot \frac{SACHMOS_j}{\sum_j SACHMOS_j}$$

where $ADLTCNT$ is the number of adults in the household and $SACHMOS_j$ is the measure of size of child j. The measure of size for a child is discussed in detail in *CHIS 2009 Methodology Series: Report 1 - Sample Design*, but we note here that within the same household children age 0 to 5 years have a measure of size twice that of children age 6 to 11 years. If the sampled adult has a spouse/partner living in the household and the spouse/partner of the sampled adult is the parent or legal guardian of the sampled child then

$$CHPROB_j = \frac{1}{ADLTCNT} \left(\frac{SACHMOS_j}{\sum_j SACHMOS_j} + \frac{SACHMOS_j}{\sum_k SPCHMOS_k} \right)$$

[10] If the spouse/partner of the sampled adult is living in the household.

where $ADLTCNT$ and $SACHMOS_j$ are defined as before and $SPCHMOS_k$ the measure of size child k associated with the spouse/partner of the sample adult. The number of sampled children and sum of the initial weights are in Table B-3 (rows 1.1 through 1.3).

5.3 Other Child Weighting Adjustments

Adjustments were made to the child weights for extended interview nonresponse, telephone service poststratification, composite weight, trimming influential weights, and raking to control totals. The child nonresponse adjustment is the same as the adult nonresponse adjustment described in Section 4.2, except the adjustment cells are defined differently. We initially created child nonresponse adjustment cells using three variables: household mailable status, sex of child, and age group (0-3, 4-7, and 8-11 years old) within sampling stratum. Since a majority of these cells had fewer than 30 respondents, we collapsed cells to increase the number of respondents in each cell. To do this we inspected adjustment factors separately by mailable status, sex, and age group at the state level to determine the variables with the most variable response rates. Using these results, for two sampling strata mailable status, sex and age group were used, and for the rest of the sampling strata the cells were defined by sampling stratum, sex of child, and age group. Any cells still containing fewer than 30 respondents were collapsed over age group. The two strata with smallest sample sizes were collapsed across both sex and age group. Table B-3 (rows 2.1 through 2.3) shows the number of sample records and sum of weights before and after the nonresponse adjustments.

The next step in weighting was to combine the landline and cell samples. As for the adult weights, child and adolescent interviews from the landline/surname and cell phone samples were poststratified separately to control totals defined by telephone service (i.e., persons in landline only households, persons in cell phone only households, and persons in households with both services). Appendix B, Table B-6 (rows 1.1 through 1.3) shows the sum of weights before and after this adjustment.

In the next step, the landline and cell phone sample were combined using a composite factor. We used the same composite factor $\lambda = 0.9$ as in the adult sample to reduce the bias of estimates computed using both samples. Appendix B, Table B-6 (row 2.1) shows the sum of weights after this adjustment.

The next step was to identify and trim large child weights. The process used for trimming the adult weights was applied to the child weights. As a result of applying the procedures, we identified and trimmed a total of 71 child weights in CHIS 2009. The trimming factors range from 0.0957 to 0.9968. Table B-6 (rows 2.1 and 3.1 through 3.3) shows the distribution of trimmed weights by self-reported stratum and the sum of the weights before and after applying the trimming factors.

The trimmed child weights were then raked to population control totals to produce estimates consistent with the California Department of Finance 2009 population estimates. See Chapter 7 for the specific controls used. The expression for the raking adjustment is the same as that for adult weights described in Section 4.4. Table B-6 (rows 3.3 and 4.1 through 4.5) shows the counts and sum of weights before and after the raking adjustments.

6. ADOLESCENT WEIGHTING

In CHIS 2009, adolescents were sampled and responded to the interview for themselves after parental permission to conduct the interview. In this section, we describe the creation of analytic weights for the adolescent interview. The steps for the adolescent weighting are similar to those for children described in the previous chapter. The format of this chapter follows that for the child weighting, with the creation of the adolescent initial weights and the adjustments for nonresponse, telephone service poststratification, composite weight, trimming, and raking.

6.1 Initial Adolescent Weights

The procedures for creating the adolescent weights are the same as those for creating the child weights described in Chapter 5. As with the child weighting, the initial weights for the adolescents incorporate the probability of sampling the adult and the probability of sampling an adolescent among the adolescents associated with the sampled adult. The initial weight, $TNA0W_i$, is

$$TNA0W_j = \frac{1}{TNPROB_j} HHA10W_i$$

where $HHA10W_i$ is defined in Chapter 5, and $TNPROB_i$ is computed the same as $CHPROB_i$. However, the measure of size is unity for all adolescents regardless of their age. Table B-4 shows the number of sampled adolescents (row 1.1) and the sum of the initial adolescent weights (row 1.2).

6.2 Other Adolescent Weighting Adjustments

The adolescent initial weight was then adjusted for nonresponse the same way the adult and child initial weights were adjusted. Note that nonresponse for the adolescent interview includes failure to obtain permission for the interview, as well as failure to interview the adolescent once permission was obtained. Table B-4 in Appendix B shows the nonresponse-adjusted adolescent weight. Initially the adolescent nonresponse adjustment cells were created using household mailable status, sex of the adolescent, and age group (12-14 and 15-17 years old) within sampling strata. We inspected response rates separately by the three variables at the state level.

After reviewing these rates, we created cells using sampling stratum, mailable status, sex and age group. Cells containing fewer than 30 respondents were collapsed across age group first and then across mailable status and sex if necessary.

The next step in weighting was to combine the landline and cell samples. As in the adult weights, child and adolescent interviews from the landline/surname and cell phone samples were poststratified separately to control totals defined by telephone service (i.e., persons in landline only households, persons in cell phone only households, and persons in households with both services). Appendix B, Table B-7 (rows 1.1 through 1.3) shows the sum of weights before and after this adjustment.

After poststratification, the landline and cell phone sample were combined using a composite factor. We used the same composite factor $\lambda = 0.9$ as in the adult sample to reduce the bias of estimates computed using both samples. Appendix B, Table B-7 (row 2.1) shows the sum of weights after this adjustment.

After the creating the composite weight, 48 influential weights were identified and trimmed, with factors ranging from 0.1359 to 0.9704. Table B-7 (rows 2.1 and 3.1 through 3.3) gives the trimmed weights by self-reported stratum and the sum of the weights before and after applying the trimming factors to the adolescent weights

In the last steps, the adolescent weights were raked to California DOF 2009 Population Estimates. See Chapter 7 for details on the control totals. The expression for the raking adjustment is the same as in the raking of the adult weights and the child weights. Table B-7 (rows 3.3 and 4.2) show the sum of weights before and after raking.

7. RAKING AND CONTROL TOTALS

This chapter describes the raking procedure and the development of control totals for CHIS 2009. The first section gives a general overview of raking and why this procedure was used in this and previous cycles of CHIS. The second section describes the 11 dimensions used to rake the weights. Eight of the dimensions are defined by demographic variables and two are defined by socio-economic variables. The 11^{th} dimension was created to reduce the bias associated with households without a landline telephone. The third section describes how the control totals were derived from the California DOF files.

7.1 Raking Procedure

Raking is an adjustment procedure in which estimates are controlled to marginal population totals. The main advantage of raking over poststratification is that raking allows the use of more auxiliary information. A limitation in poststratification is that each unit falls into only one adjustment cell and the number of respondents in a cell could be too small. With raking, the cell size is based on the distribution of each raking dimension. For example, with poststratification, only some cross-classified age/race/sex categories could be used in the adjustments, whereas with raking the full cross-classification is not needed, and important geographic data such as county can be included as dimensions. Raking may be thought of as a multidimensional poststratification procedure because the weights are basically poststratified to one set (a dimension) of control totals, then these adjusted weights are poststratified to another dimension. After all dimensions are adjusted, the process is iterated until the control totals for all the dimensions are simultaneously satisfied within a specified tolerance. Raking was also used in previous cycles of CHIS. Below, we describe the procedure in more detail. Brackstone and Rao (1979); Deville and Särndal (1992); and Kalton and Flores Cervantes (2003) also describe raking.

For simplicity, consider two auxiliary variables (or dimensions) with C and D classes, respectively. If we cross-classify the two variables into C*D cells and the sample counts in some cells are small, then it is likely that the poststratified estimates may be unstable unless the cells in the cross-tabulation are collapsed. With the 11 dimensions used in CHIS 2009, the potential collapsing would be very extensive.

An alternative approach is to rake the weights to the marginal totals of the variables. The raking-adjusted estimator is design-unbiased in large samples and is very efficient in reducing the variance of the estimates if the estimates in the cross-tabulation are consistent with a model that ignores the interactions between variables. Collapsing is sometimes required with raking, but it is not as extensive as with poststratification.

The raked weights can be written as $\tilde{w}_{cd,i} = w_{cd}\hat{\alpha}_c\hat{\beta}_d$, where w_{cd} is the pre-raked weight of an observation in cell (*c*, *d*) of the cross-tabulation, $\hat{\alpha}_c$ is the effect of the first variable, and $\hat{\beta}_d$ is the effect of the second variable. Note that in this formulation there is no interaction effect; the weights are determined by the marginal distributions of the control variables. As a result, the sample sizes of the marginal distributions are the important determinants of the stability of the weighting procedure, not the cells formed by the cross-classification of the variables. Deficient cells (cells with small sample sizes) are thus defined in terms of the sample sizes of the marginal distributions, not of the cross-classified cells.

7.2 Raking Dimensions

The 11 dimensions used in CHIS 2009 are shown in Table 7-1. The first eight dimensions in Table 7-1 were created by combining demographic variables (age, sex, race, and ethnicity) and different geographic areas (city, county, region or group of counties, and state). The 9th, 10th, and 11th dimensions use additional variables. The 11th dimension was specifically created to adjust the weights for households without a landline telephone. This dimension was also used to create the combined landline/list/cell sample weights. Section 7.3 describes this adjustment and the variables used to create the levels for this dimension. The raking dimensions for CHIS 2009 are similar to those used in previous CHIS cycles, which all also include 11 dimensions.

Table 7-1. Definitions of the dimensions used in raking

Dimension	Level	Description	Categories	
1	Stratum (collapsed where necessary)	Age groups (3) x Sex (2)	11	Under 12 years, male
			12	Under 12 years, female
			21	12 to 17 years, male
			22	12 to 17 years, female
			31	18 years or older, male
			32	18 years or older, female

See note at end of table.

Table 7-1. Definitions of the dimensions used in raking (continued)

Dimension	Level	Description	Categories	
2	Stratum (collapsed where necessary)	Age groups (9)	1	Under 6 years
			2	6 to 11 years
			3	12 to 17 years
			4	18 to 24 years
			5	25 to 29 years
			6	30 to 39 years
			7	40 to 49 years
			8	50 to 64 years
			9	65 years or older
3	State	Age groups (13) x Sex (2)	11	Under 4 years, male
			12	Under 4 years, female
			21	4 to 7 years, male
			22	4 to 7 years, female
			31	8 to 11 years, male
			32	8 to 11 years, female
			41	12 to 14 years, male
			42	12 to 14 years, female
			51	15 to 17 years, male
			52	15 to 17 years, female
			61	18 to 24 years, male
			62	18 to 24 years, female
			71	25 to 30 years, male
			72	25 to 30 years, female
			81	31 to 37 years, male
			82	31 to 37 years, female
			91	38 to 45 years, male
			92	38 to 45 years, female
			101	46 to 53 years, male
			102	46 to 53 years, female
			111	54 to 64 years, male
			112	54 to 64 years, female
			121	65 to 77 years, male
			122	65 to 77 years, female
			131	78 years or older, male
			132	78 years or older, female
4	SPAs in Los Angeles Co., HRs in San Diego Co., Remainder of CA	SPAs (8), HRs (6), Remainder of CA (1)	1	SPA 1 – Antelope Valley
			2	SPA 2 – San Fernando
			3	SPA 3 – San Gabriel
			4	SPA 4 – Metro
			5	SPA 5 – West
			6	SPA 6 – South
			7	SPA 7 – East
			8	SPA 8 – South Bay
			9	HR 1 – North Coastal
			10	HR 2 – North Central
			11	HR 3 – Central
			12	HR 4 – South
			13	HR 5 – East
			14	HR 6 – North Inland
			15	Remainder of CA

See note at end of table.

Table 7-1. Definitions of the dimensions used in raking (continued)

Dimension	Level	Description		Categories
5	Region (collapsed where necessary)	Race/ethnicity (7)	1	Latino
			2	Non-Latino White
			3	Non-Latino African American
			4	Non-Latino American Indian
			5	Non-Latino Asian
			6	Non-Latino Native Hawaiian
			7	Non-Latino Two or more races
6	State	Gender (2) x Race/ethnicity (7) x Age groups (4) (collapsed where necessary)	111	Male, Latino, under 12 years
			112	Male, Latino, 12 to 17 years
			113	Male, Latino, 18 to 64 years
			114	Male, Latino, 65 years or older
			121	Male, Non-Latino White, under 12 years
			122	Male, Non-Latino White, 12 to 17 years
			123	Male, Non-Latino White, 18 to 64 years
			124	Male, Non-Latino White, 65 years or older
			131	Male, Non-Latino African American, under 12 years
			132	Male, Non-Latino African American, 12 to 17 years
			133	Male, Non-Latino African American, 18 to 64 years
			134	Male, Non-Latino African American, 65 years or older
			141	Male, Non-Latino American Indian, under 12 years
			142	Male, Non-Latino American Indian, 12 to 17 years
			143	Male, Non-Latino American Indian, 18 to 64 years
			144	Male, Non-Latino American Indian, 65 years or older
			151	Male, Non-Latino Asian, under 12 years
			152	Male, Non-Latino Asian, 12 to 17 years
			153	Male, Non-Latino Asian, 18 to 64 years
			154	Male, Non-Latino Asian, 65 years or older
			161	Male, Non-Latino Native Hawaiian, under 12 years
			162	Male, Non-Latino Native Hawaiian, 12 to 17 years
			163	Male, Non-Latino Native Hawaiian, 18 to 64 years
			164	Male, Non-Latino Native Hawaiian, 65 years or older
			171	Male, Non-Latino Two or more races, under 12 years
			172	Male, Non-Latino Two or more races, 12 to 17 years
			173	Male, Non-Latino Two or more races, 18 to 64 years
			174	Male, Non-Latino Two or more races, 65 years or older
			211	Female, Latino, under 12 years
			212	Female, Latino, 12 to 17 years
			213	Female, Latino, 18 to 64 years
			214	Female, Latino, 65 years or older
			221	Female, Non-Latino White, under 12 years
			222	Female, Non-Latino White, 12 to 17 years
			223	Female, Non-Latino White, 18 to 64 years
			224	Female, Non-Latino White, 65 years or older
			231	Female, Non-Latino African American, under 12 years
			232	Female, Non-Latino African American, 12 to 17 years

See note at end of table.

Table 7-1. Definitions of the dimensions used in raking (continued)

Dimension	Level	Description		Categories
			233	Female, Non-Latino African American, 18 to 64 years
			234	Female, Non-Latino African American, 65 years or older
			241	Female, Non-Latino American Indian, under 12 years
			242	Female, Non-Latino American Indian, 12 to 17 years
			243	Female, Non-Latino American Indian, 18 to 64 years
			244	Female, Non-Latino American Indian, 65 years or older
			251	Female, Non-Latino Asian, under 12 years
			252	Female, Non-Latino Asian, 12 to 17 years
			253	Female, Non-Latino Asian, 18 to 64 years
			254	Female, Non-Latino Asian, 65 years or older
			261	Female, Non-Latino Native Hawaiian, under 12 years
			262	Female, Non-Latino Native Hawaiian, 12 to 17 years
			263	Female, Non-Latino Native Hawaiian, 18 to 64 years
			264	Female, Non-Latino Native Hawaiian, 65 years or older
			271	Female, Non-Latino Two or more races, under 12 years
			272	Female, Non-Latino Two or more races, 12 to 17 years
			273	Female, Non-Latino Two or more races, 18 to 64 years
			274	Female, Non-Latino Two or more races, 65 years or older
7	State	Asian groups (5) x Age groups (4) (collapsed where necessary)	11	Non-Latino Chinese only, under 12 years
			12	Non-Latino Chinese only, 12 to 17 years
			13	Non-Latino Chinese only, 18 to 64 years
			14	Non-Latino Chinese only, 65 years or older
			21	Non-Latino Korean only, under 12 years
			22	Non-Latino Korean only, 12 to 17 years
			23	Non-Latino Korean only, 18 to 64 years
			24	Non-Latino Korean only, 65 years or older
			31	Non-Latino Filipino only, under 12 years
			32	Non-Latino Filipino only, 12 to 17 years
			33	Non-Latino Filipino only, 18 to 64 years
			34	Non-Latino Filipino only, 65 years or older
			41	Non-Latino Vietnamese only, under 12 years
			42	Non-Latino Vietnamese only, 12 to 17 years
			43	Non-Latino Vietnamese only, 18 to 64 years
			44	Non-Latino Vietnamese only, 65 years or older
			51	Other or non-Asian only, under 12 years
			52	Other or non-Asian only, 12 to 17 years
			53	Other or non-Asian only, 18 to 64 years
			54	Other or non-Asian only, 65 years or older

See note at end of table.

Table 7-1. Definitions of the dimensions used in raking (continued)

Dimension	Level	Description	Categories	
8	Stratum (collapsed where necessary)	Race/ethnicity (3) x Age groups (4)	11	Latino, under 12 years
			12	Latino, 12 to 17 years
			13	Latino, 18 to 64 years
			14	Latino, 65 years or older
			21	Non-Latino White, under 12 years
			22	Non-Latino White, 12 to 17 years
			23	Non-Latino White, 18 to 64 years
			24	Non-Latino White, 65 years or older
			31	Non-Latino Non-White, under 12 years
			32	Non-Latino Non-White, 12 to 17 years
			33	Non-Latino Non-White, 18 to 64 years
			34	Non-Latino Non-White, 65 years or older
9	Region (collapsed where necessary)	Education (4)	1	Not applicable (age < 18 years)
			2	Less than High School
			3	High School grad or GED recipient
			4	At least some college
10	Region (collapsed where necessary)	# Adults in HH (3)	1	0 or 1 adult,
			2	2 adults,
			3	3 or more adults
11	Region(collapsed where necessary)	Nonlandline telephone dimension		See Table 7-3

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

Before raking the weights, dimensions with levels or cells with fewer than 50 respondents were collapsed with "adjacent" cells. In dimensions 1, 2, 5, and 8 the collapsed cells were created within the geographic regions shown in Table 7-2. As Dimension 5 was defined at the region level, some cells were collapsed across regions if the regions did not contain enough respondents. Dimensions 3, 6, and 7 were defined at the state level because there were too few respondents in many of the cells at lower geographic levels. Dimensions 9, 10 and 11 were defined at the region level because the control totals needed to create these cells (education and type of household defined by number of adults in the household) were not available at the county level. When collapsing the cells, we ensured that there was at least one cell or a group of cells within each self-reported stratum. In this way, the raked weights summed to the total number of persons in each stratum.

Dimensions 9, 10 and 11 were defined at the region level because the control totals needed to create these cells (education and type of household defined by number of adults in the household) were not available at the county level. When collapsing the cells, we ensured that there

was at least one cell or a group of cells within each self-reported region. In this way, the raked weights summed to the total number of persons in each stratum.

Table 7-2. Regions in California

Region	Counties
Northern & Sierra Counties	Butte, Shasta, Humboldt, Lake, Mendocino, Yuba, Nevada, Sutter, Colusa-Glenn-Tehama, Del Norte-Lassen-Modoc-Plumas-Sierra-Siskiyou-Trinity, Alpine-Amador-Calaveras-Inyo-Mariposa-Mono-Tuolumne
Greater Bay Area	Santa Clara, Alameda, Contra Costa, San Francisco, San Mateo, Sonoma, Solano, Marin, Napa
Sacramento Area	Sacramento, Placer, Yolo, El Dorado
San Joaquin Valley	Fresno, Kern, San Joaquin, Stanislaus, Tulare, Merced, Kings, Madera
Central Coast	Ventura, Santa Barbara, Santa Cruz, San Luis Obispo, Monterey, San Benito
Los Angeles	Los Angeles
Other Southern California	San Diego, Orange, San Bernardino, Riverside, Imperial

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

7.3 Nonlandline Telephone Adjustments

CHIS 2009 included a cell phone sample to address the problem of traditional telephone samples where households without a landline telephone —including households with only cellular telephones— do not have a chance of being selected. The potential for bias from this undercoverage is related to the percentage of households without landline telephones and the differences in characteristics of the persons residing in households with landlines and those without.

CHIS 2005 was the first cycle to include a nonlandline adjustment focused on reducing the potential bias introduced by exclusion of wireless-only households from the survey. Like the nontelephone adjustment since CHIS 2001 and CHIS 2003, this adjustment was included as an additional raking dimension at the person level (dimension 11). The adjustment could not be done at the household level because it required data from the adult extended interview.

To achieve consistency across CHIS cycles, the weights include the same non-landline raking dimension implemented since CHIS 2005. The main reason for the inclusion in raking is that this dimension does not control the total number of adults in cell-only households but the total of adults with characteristics related to this population.

The goal of the nonlandline adjustment is to adjust the weights of adults in a weighting cell with a similar propensity of having a landline telephone. The variables used to create the raking cells were those identified by Blumberg et al. (2006) as good predictors of whether a household has a landline telephone. The control totals were derived for the same cells using the 2009 California Department of Finance (DOF) Population Estimates and the 2008 ACS public use micro data file (ACS-PUMS). Table 7-3 shows the definition of the 16 cells of dimension 11 used for the nontelephone adjustment in CHIS 2009. In CHIS 2009, this raking dimension was not created at the state level but included separate cells for Los Angeles County and the rest of California.

Table 7-3. Dimension 11, nonlandline telephone adjustment cell definition for CHIS 2009

Dimension 11 levels	Stratum	Household tenure	Age in years	Educational attainment	Number of adults in the household
1	Los Angeles	Own	0 to 17	NA	0 or 1
2		Rent	0 to 17	NA	0 or 1
3		Own	0 to 17	NA	2 or more
4		Rent	0 to 17	NA	2 or more
5		Own	18 to 30	Up to high school	NA
6		Own	31 to 64	Up to high school	NA
7		Own	65 and older	Up to high school	NA
8		Own	18 to 30	Greater than high school	NA
9		Own	31 to 64	Greater than high school	NA
10		Own	65 and older	Greater than high school	NA
11		Rent	18 to 34	Up to high school	NA
12		Rent	35 and older	Up to high school	0 or 1
13		Rent	35 and older	Up to high school	2 or more
14		Rent	18 to 34	Greater than high school	NA
15		Rent	35 and older	Greater than high school	0 or 1
16		Rent	35 and older	Greater than high school	2 or more
101	Remainder of California	Own	0 to 17	NA	0 or 1
102		Rent	0 to 17	NA	0 or 1
103		Own	0 to 17	NA	2 or more
104		Rent	0 to 17	NA	2 or more
105		Own	18 to 30	Up to high school	NA
106		Own	31 to 64	Up to high school	NA
107		Own	65 and older	Up to high school	NA
108		Own	18 to 30	Greater than high school	NA
109		Own	31 to 64	Greater than high school	NA
110		Own	65 and older	Greater than high school	NA
111		Rent	18 to 34	Up to high school	NA
112		Rent	35 and older	Up to high school	0 or 1
113		Rent	35 and older	Up to high school	2 or more
114		Rent	18 to 34	Greater than high school	NA
115		Rent	35 and older	Greater than high school	0 or 1
116		Rent	35 and older	Greater than high school	2 or more

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

7.4 Raking Factors

Table 7-4 shows the overall and relative raking adjustment factors for the adult, child, and adolescent weights for the landline/list sample. The overall adjustment factors were computed as the ratio of the control total to the sum of weights before raking. The factors in the table cannot be compared to those from previous cycles of CHIS because the 2009 weights include a telephone service adjustment (Section 4.3, 5.3, and 6.2) not used in previous years. As a result, the raking factors cannot be used as a measure of person-level undercoverage at the state level. Nevertheless, they may be used as an indicator of which groups were harder to reach, or were less likely to complete the interview. Larger adjustment factors suggest relative undercoverage and smaller factors relative overcoverage.

Table 7-4. Overall adjustment raking factors for adult, child, and adolescent interviews by sample characteristics

Characteristic	Adult	Child	Adolescent
Sex			
Male	1.063	0.888	1.028
Female	0.935	1.126	0.968
Age group			
Under 5 years		1.071	
6 – 11 years		0.919	
12 – 17 years			0.997
18-24 years	1.303		
25-29 years	1.249		
30-39 years	1.348		
40-49 years	1.150		
50-64 years	0.802		
65 years and over	0.681		
Race/Ethnicity[a]			
Latino	1.312	0.987	1.063
Non-Latino			
White alone	0.797	0.928	0.847
African American alone	1.127	1.289	1.159
American Indian/Alaska Native alone	0.852	1.035	0.827
Asian alone	1.434	1.278	1.562
Native Hawaiian and Other Pacific Islander alone	2.377	0.930	1.031
Two or more races	0.646	0.767	0.696
Non-Latino Asian ethnic groups			
Chinese only	1.115	0.759	0.914
Korean only	1.384	1.061	2.101
Filipino only	2.051	3.509	5.379
Vietnamese only	2.288	1.863	3.478
Educational Attainment			
Not applicable (age < 18 years)		0.990	0.997
Less than High School,	1.505		
High School grad or GED recipient,	1.148		
Some college	0.859		
College degree or above	0.862		

Table 7-4. Overall adjustment raking factors for adult, child, and adolescent interviews by sample characteristics (continued)

Characteristic	Adult	Child	Adolescent
Household Tenure [a]			
Owner	0.977	1.088	1.048
Renter	1.022	0.901	0.925
Number of adults in the household [b]			
One	0.714	1.141	1.230
Two	0.964	1.079	1.003
Three or more	1.225	0.769	0.907
Number of children in the household [b]			
None	0.930	.	1.048
One	1.200	0.991	0.955
Two or more	1.296	0.990	0.898
Number of adolescents in the household [b]			
None	0.969	1.006	.
One	1.151	1.003	1.044
Two or more	1.158	0.833	0.948

[a] OMB race ethnicity

[b] Person level estimate by type of household

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

Table 7-4 shows that for adults, the adjustment factor for males is larger than for females, which is common in household surveys. The factors also suggest undercoverage of younger adults (under 50 years old), and adults who do not own their home, adults in households with three or more adults, adults in households with at least one child or adolescent.

One large adjustment factor is for persons who self-reported as having less than a high school education (1.505). The factors for the Latino, non-Latino Asian, non-Latino African American and non-Latino Native Hawaiian and Other Pacific Islander groups are also all larger suggesting potential undercoverage. Other factors worth noting include those for persons who self-reported as being two or more races. The small factors for these race groups suggests the CHIS 2009 estimates of persons of two or more races before raking are much higher than the corresponding DOF 2009 totals.

7.5 Sources Used to Produce the Control Totals for CHIS 2009

In all cycles of CHIS considerable thought was given to the choice of data for the primary source of the control totals. In CHIS 2001, Census 2000 data were originally used because those data were recently compiled. However, as the cycles of CHIS move away from 2000, census data do not reflect the current population as well. During the CHIS 2003 cycle, several sources for control totals were examined, including Census 2000 files (U.S. Census Bureau, 2001), the 2002 American Community Survey (ACS) (U.S. Census Bureau, 2003), and the 2003 California DOF Population Projections (State of California, Department of Finance, 2004). The DOF projections were settled on as the primary source for control totals in 2003 with the hope that they could also be used for future cycles of CHIS. The DOF provides Population Projections at the county level by race, ethnicity, gender and single age for each year from 2000 to 2050. The DOF also provides Population Estimates (State of California, Department of Finance, 2006a). The estimates are updated projections based on current birth and death data. The difference between the DOF projections and estimates is that the former are produced prior to the projected year and the latter after the estimated year. Therefore, the distributions of the DOF Population Estimates are more representative of the population.

Based on discussions with UCLA, the 2009 California DOF Population Projections poststratified to 2009 DOF Population Estimates were used as the primary source of control totals for the demographic control totals (i.e., raking dimensions defined by gender, race, ethnicity, age, and stratum) for CHIS 2009.

As in previous cycles of CHIS, the population totals had to be adjusted to remove the population living in group quarters who was not eligible for the survey. The 2000 Census files were used to compute the proportion of persons living in group quarters. The 2008 ACS public use files were also used as a source for educational attainment, household tenure, and household composition.

7.5.1 California Department of Finance Population Predictions

The DOF population projections are provided at the county level by gender, race/ethnicity and single age for each year as indicated in Table 7-5. The DOF population projections used the 2000 Census counts not adjusted for the Census 2000 undercount as the baseline. A baseline cohort-component method was used to project population estimates based on fertility/mortality rates and life expectancy by different race-ethnic groups and age cohorts. Special populations (prisons, colleges, and military installations) that have very different demographic and behavioral characteristics from the rest of the population were removed from the baseline and projected separately. However, the DOF held most of the special populations only at the year 2000 level. This factor played an important role in the assumptions made when removing the population living in group quarters from the control totals in CHIS 2009.

Table 7-5. Definition of counts available in the 2009 California DOF population projections files*

Variable	Available counts
Age groups (101)	Age 0 Age 1 . . . Age 100 or more
Sex (2)	Male Female
Race-ethnicity (6)	Latino, any race Non-Latino White alone Non-Latino African American alone Non-Latino American Indian/Alaska Native alone Non-Latino Asian alone Non-Latino Native Hawaiian and Other Pacific Islander alone Non-Latino Two or more races

* Available at the county level

Source: State of California, Department of Finance.

In previous years, the DOF provided an additional file with population counts for Latinos by race. This file was used to create the population counts by the full cross-tabulation age × sex × ethnicity × race. However, beginning in 2006, the DOF stopped providing this file. We used the corresponding file for 2005 to compute the proportion of race groups within Latinos and apply them to the total Latino population in the 2009 DOF file.

The main disadvantage of the DOF projections is the race categorization. The DOF population estimates follow the U.S. Office of Management and Budget (OMB) race definition known as "modified" race with no separate population counts for "other" race. The DOF estimates comply with the OMB 1997 revised standards for collection, tabulation, and presentation of federal data on race and ethnicity (Office of Management and Budget, 1997). The revised OMB standards identify only five main racial categories (White, Black or African American, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander) and combinations of these categories. In CHIS, respondents who could not identify themselves as any of the five OMB race categories could answer with a sixth category, "some other race," consistent with the 2000 Census data collection method. Recoding of "other race" for CHIS 2009 largely followed Census procedures (see *CHIS 2009 Methodology Series: Report 3 - Data Preparation*).

In order to use the DOF estimates, any sampled person who reported "other race" (alone or in combination with another race) had to be recoded into the OMB categories. This would have required the imputation of an OMB race category for 9,822 persons (14.3 percent of the CHIS 2009 sample) who self-reported "other race" only. As an alternative, a variable combining ethnicity with the OMB race that reduced the number of imputations was proposed and approved by UCLA. The recoding includes an additional level that arranges Latinos of any race into one group as shown in Table 7-6. Because most respondents who reported "other race" only were Latinos, the number of imputed records was reduced significantly to 108 persons (0.16 percent) who reported a non-Latino other race. The advantage of this additional variable is that it matches the categories of the population projections available in the DOF files. See Section 8.4 for additional details for the creation and imputation of this variable, OMBSRREO.

Table 7-6. Description of the variable using the OMB race definition

OMBSRREO	Definition
1	Latino
2	Non-Latino White alone
3	Non-Latino African American alone
4	Non-Latino Asian alone
5	Non-Latino American Indian/ Alaska Native alone
6	Non-Latino Pacific Islander alone
7	Non-Latino two or more races

The DOF population estimates include the population living in group quarters. Since the target population in CHIS 2009 excludes persons in group quarters, these persons were removed from the DOF population projections. At the time control totals were being developed, the DOF did not have separate projections for the population living in group quarters. The Census 2000 files were used to estimate the proportion of persons in group quarters, and these proportions were applied to the DOF estimates.

7.5.2 Census 2000 Files

The Census files were used to derive the control totals for the dimension defined by SPAs in Los Angeles and Health Regions in San Diego County (dimension 4 in Table 7-1). The proportions of the total population in those areas were computed from the 2000 Census files. This assumes that the proportion in these areas with respect to the county did not change between 2000 and 2007[11]. The Los Angeles SPAs and San Diego Health Regions were both defined in terms of Census Tracts.

7.5.3 American Community Survey for California

The American Community Survey (ACS) is a nationwide survey that provides current and detailed demographic, social, economic, and housing data. It is a critical element in the Census Bureau's reengineered 2010 Census plan as it has replaced the decennial census long form The 2008 ACS produces population and household estimates for a limited number of characteristics at the state level and for over 800 geographical areas excluding the group quarters populations. There is a 2008 ACS public use micro data file (ACS-PUMS) for California that provides household and population estimates at the state level. For selected counties and large communities, custom tables for a subset of estimates can be downloaded from the Census website (http://www.census.gov/acs/www/). The 2008 ACS includes population estimates for 24 CHIS 2009 strata, but not for the SPAs in Los Angeles County, San Diego County Health Regions, or many small counties.

Although the 2008 ACS was not used as the primary source of population control totals in CHIS 2009, it was used to compute proportions by educational attainment and type of household (tenure and number of adults in the household) at the region level as these variables were not available in the DOF files. These proportions were applied to the 2009 DOF total population counts to derive the control totals for the raking dimensions defined by these characteristics (dimensions 9, 10, and 11 in Table 7-1). The proportions were calculated at the region level allowing use of the ACS data. Applying the 2008 factors assumed that there were no changes in the population proportions between 2008 and 2009 for these variables.

7.5.4 The National Health Interview Survey

The National Health Interview Survey (NHIS) which is one of the major data collection programs of the National Center for Health Statistics (NCHS) which is part of the Centers for Disease Control and Prevention (CDC) has been conducted since 1957. The NHIS is an in-person survey where sampling and interviewing are conducted continuously throughout the

[11] The population in group quarters was removed from these areas and the county before computing the proportions.

year. The survey collects information about household telephone and whether anyone in the household has a wireless telephone. This survey has been used to track wireless substitution in the US. We used estimates from the Early Release program from January to June 2010 to compute the percentages of adolescents, children, and adolescents by type of telephone service in the household (i.e., landline only, cell phone only, or both). Because the NHIS does not produce estimates at the state level, we use the estimated for the West region. Table 7-7 these percentages that were applied to the DOF totals to derive the control totals for poststratification.

Table 7-7. Description of the variable using the OMB race definition

Person type	Telephone service	Proportion
Adults	Landline only	0.129
	Cell phone and landline	0.646
	Cell phone only	0.225
Children and adolescents	Landline only	0.108
	Cell phone and landline	0.647
	Cell phone only	0.245

7.6 Producing the Control Totals for CHIS 2009

As with previous cycles of CHIS, the derivation of the control totals was a challenging task in 2009. It involved the selection of the sources of control totals, determining the number of dimensions, and computing the control totals. It also had an impact on the set of variables to be imputed. In CHIS 2009, there were 11 raking dimensions. Deriving the control totals for each dimension independently could lead to inconsistencies between totals across the dimensions and this would cause problems in the raking process.

To overcome these difficulties, we used the same procedure developed for CHIS 2003 in which the control totals for most of the dimensions were computed simultaneously. In this approach, a file was created with totals for all the possible combinations of the levels from most of the raking dimensions in the source files. These totals were then adjusted to remove the population living in group quarters. In the final step, the file was summarized by aggregating the totals by raking dimension. Because all totals were produced from the same file, there were no inconsistencies (the sum across dimensions was constant and the relationship between dimensions using the same variables such as age was fixed) among the dimensions. The details of this procedure are described in the following sections.

As the first step when computing control totals, the population living in group quarters was removed from the population counts. This is a straightforward process when counts of persons in group quarters for all variables and geographic levels are available. However, this information was not available in the DOF files. By assuming that the proportion of the population in group quarters did not change between 2000 and 2009, the Census 2000 SF1 file could be used to compute these proportions. This assumption is the same one used by the California DOF for its population projections.

In past cycles of CHIS, two problems occurred when computing the percentage of the population living in group quarters using the Census SF1 file. The first was the limited number of group quarter counts that can be produced from the SF1 file. Counts are available only by stratum (44) × age group 1 (3) × sex (2) × race (7) and by stratum (44) × age group 1 (3) × sex (2) × ethnicity (3) as defined in Table 7-7. Other counts included totals by stratum (44) and single year of age.[12] The file could not be used to produce population counts by single year of age by the cross-tabulation of race and ethnicity. The file with counts by single age was used to compute the population in group quarters for younger and older children and adolescents, e.g., and stratum (44) × age group 2 (5) where age group 2 is defined in Table 7-7. As in previous cycles the assumption was made that the distribution of the population in group quarters is uniform among the age groups; for example, if the percentage of persons 65 or older in group quarters is 1.56 percent, then 1.56 percent of persons 68 years old are assumed to be in group quarters.

The second problem was that the group quarter population counts from the SF1 file are defined for the seven race categories shown in Table 7-8 and not the six OMB race groups used in the DOF file (see Table 7-5). To address this problem, we assumed that the distribution of persons in group quarters by ethnicity (Latino or non-Latino) was also the same within race. For example, if 1.42 percent of the African American population is in group quarters, then 1.42 percent of both Latino African Americans and non-Latino African Americans are assumed to be in group quarters.

[12] Census 2000 totals for the population in group quarters by single age were not used in previous cycles. These counts were used for the first time in CHIS 2007 for more accurate control of the proportion in group quarters of two age groups for children (0 to 5 and 6 to 12 years old) and adolescents. The proportion of the population in group quarters is not uniform across age groups; for example, the proportion of adolescents in group quarters is much larger than the proportion of younger children.

Table 7-8. Definition of levels of variables for group quarters populations in the Census 2000 SF1 file

Characteristics	Available counts
Stratum (44)	Counties or combinations of multiple counties defined in CHIS 2009
Age group1 (3)	Less than 18 years old 18 to 64 years old 65 years old or older
Age group2 (5)	Less than 6 years old 6 to 11 years old 12 to 17 years old 18 to 64 years old 65 years old or older
Sex (2)	Male Female
Race (7)	White alone African American alone American Indian/Alaska Native alone Asian alone Native Hawaiian and Other Pacific Islander alone Other race alone Two or more races
Ethnicity(3)	Latino Non-Latino White alone Other

Source: U.S. Census Bureau, Census 2000.

Using this assumption, the percentage of the population not living in group quarters in 2000 was computed as follows. A file with 2000 population totals, T_{rc}^{2000}, was created by summarizing the 2000 SF1 into 22,176 cells denoted *rc*, where *r* denotes race and *c* is the cross-tabulation of stratum (44) × ethnicity (2) × age group (18) × gender (2). The 18 levels of age (see Table 7-9) corresponded to the cross-tabulation of the levels of age available in the DOF data files and in the definition of the raking dimensions. An advantage of summarizing the file by the levels of *c* was the smaller size of the file (i.e., the file contains population totals by the age groups rather than single age). Note that any age group, race, or ethnicity as defined in the raking dimensions could be created by combining the *c* cells.

We defined the cells *rc* as the cross-tabulation of race and the cell *c* as follows:

$$rc = \text{race}_{\overline{OMB}}(7) \times c \,,$$

where the subscript $\overline{OMB}$ refers to the non-OMB race classification that includes a category for "some other race" available in the SF1 file as shown in Table 7-7.

Table 7-9. Age levels corresponding to the cross-tabulation of the DOF data files and the definition of the raking dimensions

Age group(i)	Description
1	0 to 3 years old
2	4 to 5
3	6 to 7
4	8 to 11
5	12 to 14
6	15 to 17
7	18 to 24
8	25
9	26 to 29
10	30
11	31 to 37
12	38 to 39
13	40 to 45
14	46 to 49
15	50 to 53
16	54 to 64
17	65 to 77
18	78 plus

7.6.1 Removing the Population Living in Group Quarters

We now review how the group quarter population was removed from the DOF files. Define $T_{rc}^{2000\,\overline{GQ}}$ as the 2000 population total that excludes the population in group quarters in cell rc. The totals $T_{rc}^{2000\,\overline{GQ}}$ were computed by raking the totals T_{rc}^{2000} to three control totals for the population not living in group quarters. Let $D1_m^{2000\,\overline{GQ}}$ be the control total for the first raking dimension computed as

$$D1_m^{2000\,\overline{GQ}} = D1_m^{2000} - D1_m^{2000\,GQ},$$

where $D1_m^{2000}$ is the 2000 total population, $D1_m^{2000\,GQ}$ is the 2000 population total living in group quarters, and m is the raking cell defined as $m = \text{strata}(44) \times \text{race}_{\overline{OMB}}(7) \times \text{age group } 1(3) \times \text{sex}(2)$.

In the same way, let $D2_n^{2000\,\overline{GQ}}$ be the control total for the second raking dimension for cell n, where n is defined as $n = strata\ (44) \times ethnicity\ (3) \times age\ group\ 1(3) \times sex\ (2)$ as in the SF1. Let $D3_p^{2000\,\overline{GQ}}$ be the control total for the third raking dimension for cell p, where p is defined as $n = strata\ (44) \times age\ group\ 2\,(5)$ as in the SF1.

Note that $D1_m^{2000\,GQ}$, $D2_n^{2000\,GQ}$, and $D3_p^{2000\,GQ}$ are the 2000 population totals living in group quarters available in the SF1 file. By using raking we ensured that all totals, $T_{rc}^{2000\,\overline{GQ}}$, were consistent and they summed to the control totals.

After raking, the proportion of the 2000 population not living in group quarters in cell *rc* was computed as

$$p_{rc}^{2000\,\overline{GQ}} = \frac{T_{rc}^{2000\,\overline{GQ}}}{T_{rc}^{2000}}.$$

Assuming that the proportion of the population not living in group quarters did not change between 2000 and 2009 within cell *rc*, the proportion $p_{rc}^{2000\,\overline{GQ}}$ could be used to compute $T_{rc}^{2009\,\overline{GQ}}$ defined as the 2009 total population not living in group in cell *rc*, as

$$T_{rc}^{2009\,\overline{GQ}} = p_{rc}^{2000\,\overline{GQ}} * T_{rc}^{2009},$$

where T_{rc}^{2009} is the 2009 total population from the 2009 California DOF file in cell *rc*. However, T_{rc}^{2009} could not be computed using the DOF file due to differences in race categorization between the SF1 and the DOF projection. Instead, the 2009 population estimates, $T_{sc}^{2009\,OMB}$, were available in the DOF file for 19,008 cells (labeled *sc*) defined using the OMB race categories. The cells *sc* were defined by the cross-tabulation of $sc = \text{race}_{OMB}(6) \times c$, where the subscript *OMB* refers to the OMB race groups that exclude the "some other race" category as shown in Table 7-10, and *c* is defined as before.

Table 7-10. OMB race categories available in the California DOF files

$\text{race}_{\text{OMB}}(s)$	Description
1-W	OMB White alone
2-AA	OMB Black or African American alone
3-AI	OMB American Indian or Alaska Native alone
4-AS	OMB Asian alone
5-PI	OMB Pacific Islander Native Hawaiian alone
6-TM	OMB Two or more races

In order to examine the relationship between the totals T_{sc}^{OMB} and T_{rc}, consider the following summation:

$$T_c^{OMB} = \sum_s T_{sc}^{OMB} = T_{Wc}^{OMB} + T_{AAc}^{OMB} + T_{AIc}^{OMB} + T_{ASc}^{OMB} + T_{PIc}^{OMB} + T_{TMc}^{OMB}.$$

In the same way, the total population in a cell *c* can be represented by non-OMB race groups as

$$T_c = \sum_r T_{rc} = T_{Wc} + T_{AAc} + T_{AIc} + T_{ASc} + T_{PIc} + T_{Oc} + T_{TMc} .$$

The assignment of OMB race was done within cell *c*; in other words, the total population in the cell *c* stays constant. That is

$$T_c = \sum_r T_{ri} = T_c^{OMB} = \sum_s T_{sc}^{OMB} .$$

When assigning an OMB race value, persons who reported "some other race" alone were assigned one of the OMB race categories. Persons who reported two races, one being "other race," kept the OMB race category but dropped "other race." In other words, they were assigned a single OMB race. Persons who reported more than two races, one of these being "other race," were still considered as having multiple races (the "other race" removed).

In order to illustrate the reallocation, consider the Asian group (ignoring the stratum, age group, sex, and ethnicity components of the cell),

$$T_{ASc}^{OMB} = T_{ASc} + p_{AS\_O}^{OMB} * T_{Oc} + p_{AS\_TM}^{OMB} * T_{TMc} ,$$

where

T_{ASc}^{OMB}	is the total number of Asians (OMB definition);
T_{ASc}	is the total number of Asians (non-OMB definition);
T_{TMc}	is the total number of persons with two or more races (non-OMB definition);
$p_{AS\_O}^{OMB}$	is the proportion of persons with some other race alone who were coded as Asian alone when assigning the OMB definition; and
$p_{AS\_TM}^{OMB}$	is the proportion of persons with two or more races who are coded as Asian alone when assigning the OMB definition.

In other words, the OMB Asian alone population (T_{ASc}^{OMB}) is composed of the original non-OMB Asian-alone total (T_{ASc}), the portion of the population who reported "some other race" alone that is allocated to OMB Asian ($p_{AS\_O}^{OMB} * T_{Oc}$), and the population who reported non-OMB Asian-alone and "some other race." Figure 7-1 visualizes how the OMB Asian-alone population is formed, where T_{AS}^{OMB} , T_{AS} , and T_{TM} are defined above and T_O is defined as the group who reported "other race" only (omitting the subscript *c* for convenience).

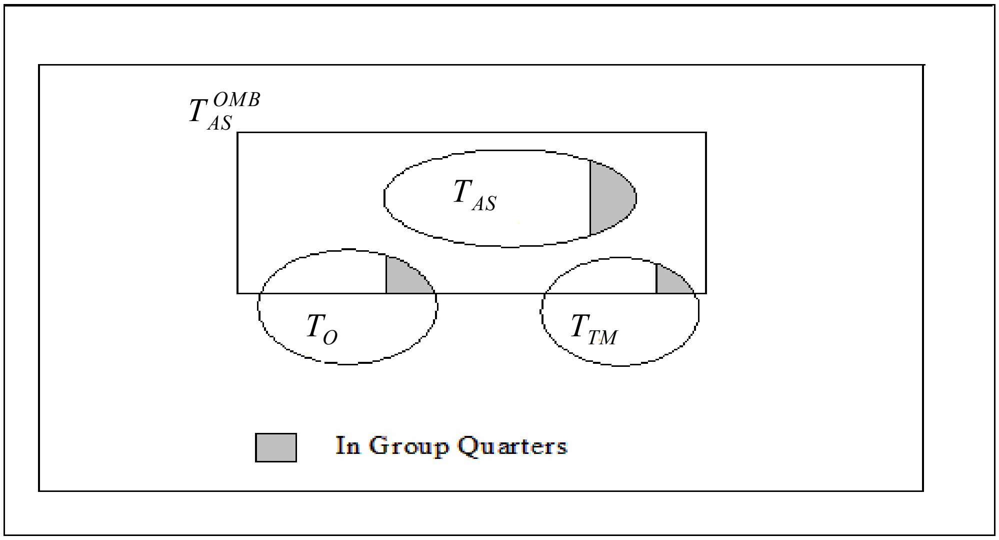


Figure 7-1. Relationship between OMB Asian alone and non-OMB groups

In this example, the proportion of the population in group quarters was known for the non-OMB Asian alone group. In order to compute the proportion of the population not in group quarters for OMB Asian alone we assumed the same proportion holds for the members that were being reclassified into the OMB race group. That is

$$\frac{T_{AS}^{OMB\,\overline{GQ}}}{T_{AS}^{OMB}} \approx \frac{T_{AS}^{\overline{GQ}}}{T_{AS}} = \frac{T_{AS}^{\overline{GQ}} + T_{O}^{\overline{GQ}} + T_{TM}^{\overline{GQ}}}{T_{AS} + T_{O} + T_{TM}},$$

only for $O \in AS$ and $TM \in AS$, (i.e., OMB race assignment to *AS*).

Generalizing these results to the other groups, the proportion of the population not in group quarters, $p_{rc}^{\overline{GQ}}$, can be computed as

$$p_{rc}^{\overline{GQ}} = \frac{T_{rc}^{\overline{GQ}}}{T_{rc}} \approx \frac{T_{sc}^{OMB\,\overline{GQ}}}{T_{sc}^{OMB}} = p_{sc}^{\overline{GQ}}$$

Under the assumption that the proportion of the population not living in group quarters did not change between 2000 and 2009 the proportion was computed as

$$p_{sc}^{2009\,\overline{GQ}} = p_{sc}^{2000\,\overline{GQ}} = \frac{T_{rc}^{2000\,\overline{GQ}}}{T_{rc}^{2000}}.$$

The proportion $p_{sc}^{2009\,\overline{GQ}}$ was used to compute the 2009 total population not living in group quarters in cell *sc*, $T_{sc}^{2009\,OMB\,\overline{GQ}}$, defined using the OMB race categories, as follows:

$$T_{sc}^{2009\,OMB\,\overline{GQ}} = p_{sc}^{2000\,\overline{GQ}} * T_{sc}^{2009\,OMB} = \frac{T_{rc}^{2000\,\overline{GQ}} * T_{sc}^{2009\,OMB}}{T_{rc}^{2000}}.$$

where T_{rc}^{2000} is computed using the SF1 file, $T_{sc}^{2009\,OMB}$ using the 2009 DOF file and $T_{rc}^{2000\,\overline{GQ}}$ is the 2000 population in cell *rc* not in group quarters, as defined earlier. The 2009 total population not living in group quarters in California is computed as

$$T^{2009\,\overline{GQ}} = \sum_{s}\sum_{c} T_{sc}^{2009\,OMB\,\overline{GQ}}.$$

Table 7-11 shows the total population in the 2009 DOF file and the estimated total (and percentage) of the population living in group quarters.

Table 7-11. Population in California in 2009 by group quarter status

Type	Population	%
In group quarters	893,743	2.34
Not in group quarters	37,361,765	97.66
Total	38,255,508	100.00

Source: California Department of Finance, 2009.

7.6.2 Computing the Control Totals

The totals $T_{sc}^{2009\,OMB\,\overline{GQ}}$ were summarized in order to compute the control totals for dimensions 1, 2, 3, 5, 6, and 8. For dimension 7, defined for Asian ethnic groups, the control totals were derived using the same totals but for Asian only. Using the Census 2000 SF1 files, we computed the percentages for the Asian groups in Table 7-12. The percentages of the Asian groups by ethnicity (Latino, non-Latino) were computed using the 2008 ACS-PUMS file. It was assumed that there were no changes in the distribution of the Asian groups between 2000 and 2008. These percentages were applied to the 2009 DOF projections.

Table 7-12. Census 2000 SF1 Asian groups

Asian Group	Description
1	Chinese alone
2	Korean alone
3	Filipino alone
4	Vietnamese alone
5	Other Asian ethnic group alone

The creation of dimension 4, defined by SPAs in Los Angeles County and Health and Human Services Agency (HHSA) Service Regions in San Diego County, used information from the Census 2000 SF1. The Los Angeles County Department of Health (LACDH) produced a listing of Census tracts by SPA. The 2000 SF1 file was used to compute the percentages of the population in the SPAs by aggregating population counts in the Census tracts. This percentage was applied to the total 2009 DOF population total (excluding group quarters) to produce the controls for dimension 4. A similar procedure was used for San Diego County Health Regions.

For dimensions 9 (adult's education attainment), 10 (number of adults in the household), and 11 (nontelephone adjustment), the percentages of the population were computed using the 2008 ACS-PUMS and then applied to the 2009 DOF population total (excluding group quarters). The underlying assumption was that there were no changes in the distribution of the population between 2008 and 2009.

8. IMPUTATION PROCEDURES

In any household survey, both unit and item nonresponse are virtually unavoidable. We have described how weighting adjustments have been used to compensate for unit nonresponse in CHIS 2009. *CHIS 2009 Methodology Series: Report 4 – Response Rates* discusses unit nonresponse in detail. This chapter focuses on item nonresponse and the imputation for missing responses of the variables used in weighting. The imputed values were needed in the last stages of the weighting process, and only interviews that were considered completed units were subject to imputation. The percentage of missing data and consequent imputation for virtually all of these items is small.

Section 8.1 describes the imputed variables and reviews the different types of imputation techniques used to fill in the missing data. The two imputation techniques employed in CHIS 2009 are random allocation and hot-deck imputation. Sections 8.2 through 8.4 discuss the imputation process for all imputed variables separately. The last section lists the geographic location variables for CHIS 2009. UCLA derived these variables after geocoding the geographic information either collected during the interview (address of respondent, nearest street intersection, self-reported county) or attached to the sample telephone (address for numbers that were mailable or ZIP Code covered by the telephone exchange).

8.1 Imputed Variables and Methods

Table 8-1 lists the variables imputed for weighting in CHIS 2009. With a few exceptions, the same variables were imputed in previous cycles of CHIS. In 2009, two additional variables were imputed in order to create a variable that defines a household by type of telephone service available to its residents. As noted above, the level of missing data is relatively small. The specific percentages of missing data are given later in the chapter. When the amount of missing data is small and assuming that the data are missing at random (i.e., the missing data have the same distribution as those with complete data within groups defined for imputation), then the bias of the estimates due to the missing data should be relatively small. The imputations may also increase the variance of the estimates, but this effect should be negligible given the low rate of missing data. A flag indicating if the response is imputed accompanies every value.

Table 8-1. Description of imputed variables

Variable name	Description	Interview items	Variable type
SRAGE	Self-reported age	AA2, CA3, TA2, KAA2	Demographic
SRSEX	Self-reported sex	AA3, CA1, TA3, KAA3	Demographic
SRTENR	Self-reported household tenure	AK25, KAK25	Socio-economic
SREDUC	Self-reported educational attainment	AH47, KAK47	Socio-economic
SRH	Self-reported Latino	AA4, CH1, TI1	Ethnicity
SRW	Self-reported white	AA5A_6, CH3_6, TI2_6	Race
SRAA	Self-reported African American	AA5A_5, CH3_5, TI2_5	Race
SRAS	Self-reported Asian	AA5A_4, CH3_4, TI2_4	Race
SRAI	Self-reported American Indian/ Alaska Native	AA5A_3, CH3_3, TI2_3	Race
SRPI	Self-reported Native Hawaiian and Other Pacific Islander	AA5A_1, AA5A_2, CH3_1, CH3_2, TI2_1, TI2_2	Race
SRO	Self-reported Other race	AA5A_7, CH3_7, TI2_7	Race
OMBSRREO	OMB self-reported race/ethnicity		Race/ Ethnicity
OMBSRASO	OMB self-reported non-Latino Asian group	AA5E_1- AA5E_18, TI7_1- TI 7_18, CH7_1- CH7_18	Race/ Ethnicity
HASCELL	Cell/Wireless telephone service in household	AM33	Telephone service
HASLANDLINE	Landline telephone service in household	AN6, REGFONE1	Telephone service

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

In CHIS 2009 and previous cycles, random allocation and hot-deck imputation were used to fill in the missing responses. The first imputation technique is a random selection from the observed distribution. This method is used only when the item missing rate is very small.

The second technique, hot-deck imputation, was used to impute race and ethnicity (including the OMB race-ethnicity variables) as well as household tenure and educational attainment in the previous cycles of CHIS. The hot-deck approach is probably the most commonly used method for assigning values for missing responses in large-scale household surveys (Sande, 1983; Ford, 1983). 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. In order to carry out hot-deck imputation for CHIS 2009, 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 characteristics. The recipient is then assigned a randomly imputed value from one of the donors in the pool. Once a donor is used, it is removed from the donor pool.

8.2 Self-Reported Sex and Age

The percentage of cases where either sex or age was missing in CHIS 2009 is very small across all samples (landline, surname list, cell phone, and area) and types of extended interviews (adult, child, and adolescent). Table 8-2 summarizes the number of cases that were imputed for sex and age. The sex of only one child, two adolescents and no adult were imputed self-reported sex. For children and adolescents, the missing data for sex was imputed randomly. A random number was generated for the three missing values.

Age was imputed in 123 cases in CHIS 2009 across all samples. A hierarchical process was followed to impute the missing self-reported age values for adults in the landline and list samples. The process used the values for self-reported age (question AA2 on the adult interview), the self-reported adult age range (question AA2A on the adult interview) asked when the adult refused to provide a specific age, the proxy-reported adult age collected during the child-first interview (question KAA2) if available, and the adult age collected during the screener interview (question ADULTAGE on the screener interview).

Table 8-2. Number and percentage of completed interviews with missing self-reported sex and age by sample type

Sample Person type	Number completed	Number missing sex	% missing sex	Number missing age	% missing age
Landline/List					
Adult	44,744	0	0.00	114	0.25
Child	8,483	1	0.01	9	0.11
Adolescent	3,191	2	0.06	0	0.00
Total	56,418	3	0.01	123	0.22
Area					
Adult	2,870	0	0.00	1	0.03
Cell Phone	462	0	0.00	0	0.00
Adult	188	0	0.00	0	0.00
Overall Total	3,520	0	0.00	1	0.03

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

The missing age for adults was imputed as follows. First, if an adult had a missing self-reported age, we checked whether the adult age was collected from a proxy adult in the child-first interview. If age was reported, this age was assigned to the sampled adult. If an age was not reported in the child-first interview, the screener age for the sampled adult was checked. If the screener age was within an age range given by the sampled adult, then screener age was used as the imputed age. If the age range was not reported, then the screener age was used. However, if the screener age was outside the reported age range, then age was randomly selected using the distribution of self-reported age within the reported age range. If no age was collected during the screener but an age range was reported, then age was randomly imputed from the distribution of self-reported age within the reported age range. If no information on age was available from any source, then age was randomly imputed using the distribution of self-reported age of all adult respondents. As an example, assume an adult respondent did not report an age but reported an age range of 40 to 44. Assume also that the proxy reported adult age in the child-first interview was 38 and the age collected in the screener interview when the adult was sampled was 38. This situation could result if the proxy misreported the sampled adult age in both the screening interview and the child-first interview. For this case, the adult age would be imputed using the distribution of the self-reported age of adults age 40 to 44. Assume that the distribution of adult age is such that 41 percent of sampled adults were age 41 or less and 62 percent were age 42 or less. If the random number assigned to the adult had a value of 0.44 then the adult's imputed age would be 42 years old.

8.3 Household Tenure and Educational Attainment

Household tenure and the adult respondent's educational attainment were used to create raking dimensions 9 and 11. Household tenure had 707 missing responses (1.18 percent), and educational attainment had 38 missing (0.06 percent).

Hot-deck imputation was used to impute missing values for these two variables. The search algorithm CHAID (Kass, 1980) was used to create the hot-deck cells using the variables available for both donors and recipients found to be good predictors. A donor was then randomly drawn from the cell and its value for the variable being imputed was assigned to the recipient. Table 8-3 shows the variables considered in CHAID to create the hot-deck cells for educational attainment and household tenure. Table 8-4 shows the distribution of the imputed cases by sample type. When calculating the percentages, the denominator for educational attainment is the number of adults in a given education category, and for tenure the denominator is all adults who own or rent.

Table 8-3. Variables used to define hot-deck cells for the imputation of education attainment and household tenure

Variable Name	Description
Educational Attainment	
SRSEX	Self-reported sex
SRRACE_O	Self-reported race
SRH	Self-reported ethnicity
SRAGE	Self-reported age
ADLTFLG	Number of adults in the household
CHLDFLG	Children present in the household
TEENFLG	Adolescents present in the household
POVERTY	Poverty
P_GRAD	Percent college graduates in exchange
P_OWN	Percent home owners in the exchange
P_BLACK	Percent African Americans in the exchange
P_HISP	Percent Latinos in the exchange
CREGION	California Regions
Household Tenure	
ADLTFLG	Number of adults in the household
CHLDFLG	Children present in the household
TEENFLG	Teens present in the household
P_GRAD	Percent college graduates in exchange
P_BLACK	Percent African Americans in the exchange
P_HISP	Percent Latinos in the exchange
P_OWN	Percent home owners in the exchange
POVERTY	Poverty
CREGION	California Regions

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

Table 8-4. Counts and percentages of imputed self-reported education attainment and household tenure

	Adult Interviews			
	Sample type			
	Landline/list		Cell phone	
	Count	Percentage	Count	Percentage
Self-reported Education Attainment				
Under 18 years of age	NA		NA	
Less than HS, 18 years of age or older	50	1.10	5	1.68
High School (or equivalent), 18 years of age or older	75	0.77	3	0.38
Some college, 18 years of age or older	55	0.46	3	0.37
BS and above, 18 years of age or older	75	0.41	3	0.31
Total	230	0.45	14	0.40
Self-reported Household Tenure				
Owner	244	0.63	14	0.78
Renter	233	1.34	10	0.58
Total	477	0.85	24	0.68

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

8.4 Self-Reported Race and Ethnicity

As described in Chapter 7, the person weights were raked to control totals from the 2009 California DOF Population Estimates. The California DOF complies with the OMB 1997 revised standards for collection, tabulation, and presentation of federal data on race and ethnicity. The revised OMB standards identify only five main racial categories and combinations of these categories. The main categories are White, Black or African American, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander. Census 2000 allowed a sixth race category ("some other race") for respondents who could not identify with any of the five OMB race categories. Because all public release files of the Census 2000 include six race categories, the Census Bureau released a special file called Modified Race Data Summary file (MRDSF) with 2000 population counts by the five OMB race categories (U.S. Census Bureau, 2002). To produce this file the Census Bureau implemented special procedures to assign and impute an OMB race to those who reported "some other race." The California DOF Estimates used the 2000 MRDSF as the baseline for the time series; as a result, the DOF Estimates include only counts by the five OMB racial categories by county.

Following a procedure similar to the Census 2000, respondents who could not identify themselves as any of the five OMB race categories could answer "some other race" in CHIS. In order to use the DOF estimates as control totals, any sampled person who reported "some other race" (alone or in combination) had to be recoded into one or more of the OMB categories. OMB race was missing 7,722 persons (12.9 percent) in CHIS 2009. After examining the procedures used by the Census to assign an OMB race, we determined that the assignment of OMB race could not be implemented using the available variables in CHIS 2009 as in Census 2000, because the number of CHIS cases in the geographic area (i.e., stratum) by Latino origin[13] cells is not large enough to guarantee a good assignment. The same situation occurred in previous cycles. To reduce the number of records to be imputed, a combined race/ethnic variable (OMBSRREO) that assigned Latinos regardless of race into one group was proposed and approved. The levels of the variable OMBSRREO are given in Table 8-5.

[13] Donors and donees must match on the specific Latino origin (Not Hispanic; Mexican; Puerto Rican, Cuban, Central American and Dominican; South American; Other Spanish).

Table 8-5. OBM race/ethnicity groups (OMBSRREO)

OMBSRREO	Description
1	Latino
2	Non-Latino White
3	Non-Latino African American
4	Non-Latino American Indian Alaskan Native
5	Non-Latino Asian
6	Non-Latino Pacific Islander Native Hawaiian
7	Non-Latino two or more races

By creating a separate group for Latinos, a valid value of OMBSRREO was missing for only 104 persons (0.16 percent) who self-reported as non-Latino and “some other race” alone[14] in 2009. The reduction in the number of cases is because most of the people who report other race were Latino. Using a variable that combined race-ethnic groups with one level of OMBSRREO for Latino eliminated the need to impute for 9,714 cases who reported Latino “other race” alone.

For continuity with the race and ethnicity variables created since 2001 (see Table 8-1), the same variables were created and imputed in 2009. We refer to these variables as the “regular” single race and ethnicity variables. The OMB race-ethnicity variable OMBSRREO was created using these regular race and ethnicity variables after imputation. Section 8.4.1 describes the imputation of the regular race and ethnicity variables while Section 8.4.2 describes the creation and imputation of the OMB race variable. Section 8.4.3 discusses the creation and imputation of self-reported Asian ethnic groups.

8.4.1 Imputation of Single Self-Reported Race and Ethnicity

While the procedures used to impute for missing values of sex and age were relatively straightforward, self-reported race and ethnicity presented a greater challenge. Different imputation methods were considered before choosing the final approach. One approach that was considered, but not adopted, was to use the self-reported race and ethnicity of a respondent to impute for any other sampled person with missing values for these items within the household. The reason this approach was not used in any cycle of CHIS is the realization that the method does not account for households with persons of more than one race and ethnicity.

[14] This includes records imputed as non-Latino “other” from the regular CHIS 2005 race imputation.

Instead a hot-deck imputation procedure was developed to deal with the diversity of race and ethnicity within households. Before describing the hot-deck approach, some special features of the race and ethnicity items are worth noting. First, although race is a series of items with subparts, the items we deal with are only those that classify a person as White, African American, Asian, American Indian/Alaska Native, Pacific Islander, or other. Also, these items are treated as either all reported or all missing. In very few cases there were missing values for one of the races but not others, but the data preparation staff was able to replace these missing values using interviewer comments. Finally, some missing values were assigned deterministically based on other items such as country of origin. These deterministic imputations were flagged like all other imputations.

Table 8-6 shows the number and percentage of cases with imputed values by type of extended interview (adult, child, and adolescent). The first columns are those cases where race is imputed, and the next set of columns is for cases where ethnicity is imputed.

Table 8-6. Number and percentage of imputed interviews with missing self-reported race and/or ethnicity

Sample type	Imputed race*		Imputed ethnicity	
Type of interview	Count	%	Count	%
Landline/list				
Adult	799	1.79	149	0.34
Child	37	0.44	37	0.44
Adolescent	177	5.57	29	0.92
Cell phone				
Adult	73	2.40	6	0.20
Child	0	0.00	0	0.00
Adolescent	9	4.52	1	0.51
Total	1095	1.83	222	0.37

* At least one value of race was imputed.
Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

The hot-deck imputations were done separately by the completed extended interview structure of the household. In general, the imputation procedure was done at the household level and handled households with the fewest missing values first and then moved to the cases with more missing values. The simplest household structure is where only an adult was interviewed (versus a household with an adult and an adolescent and/or a child). A household with only one adult with missing ethnicity was imputed before a household with only an adult that had both missing race and ethnicity.

The patterns of missing data for race and ethnicity varied by the structure of the household. For the simple case where only an adult was interviewed, the donors were selected from other adult-only households. If the adult was missing both race and ethnicity, both values were imputed from the same donor. If the adult had a reported race but was missing ethnicity, then a donor with the same race (all six race values were placed into a vector and only adults with the exact same values could be donors) was randomly selected. For an adult with reported ethnicity and missing race, the same procedure was used; only adults in adult-only households with the same value of ethnicity could be donors. Whenever possible, the donors and the recipients were from the same sampling stratum. For cases where the pool formed in this way had too few donors, sampling strata were combined based on geographic and urban status. Once a donor was used, it was removed from the pool for all future hot deck runs.

The same principles were used for more complex household structures. In these cases, some households had missing race and ethnicity for all sampled persons, while in others one or more of the sampled persons might have a reported race and ethnicity. Various combinations, such as a reported ethnicity but not race, were also encountered. Separate hot deck runs were made to accommodate all of these situations. As an illustration, consider households where an adult and a child are interviewed. Assume the adult reported non-Latino ethnicity and Asian race and the child only reported non-Latino ethnicity but no race. The pool of donors for imputing the child's race consists of households where only an adult and a child were interviewed and where the adult reported non-Latino ethnicity and Asian race and the child reported non-Latino ethnicity. The households with other combinations of persons with missing race and/or ethnicity were imputed in an similar way. Table 8-7 shows the counts and percentages of imputed values by self-reported race and ethnicity and type of extended interview (adult, child, and adolescent).

Table 8-7. Counts and percentages of imputed interviews with missing self-reported race and ethnicity by type of extended interview

	Total		Extended interview type					
			Adult		Child		Adolescent	
	Count	%	Count	%	Count	%	Count	%
Self-reported race								
White alone	448	0.75	373	0.78	28	0.31	47	1.39
African American alone	19	0.03	15	0.03	1	0.01	3	0.09
Asian alone	24	0.04	19	0.04	5	0.06	0	0.00
American Indian/ Alaska Native alone	34	0.06	20	0.04	0	0.00	14	0.41
Pacific Islander alone	3	0.01	1	0.00	0	0.00	2	0.06
Other race alone	544	0.91	429	0.90	3	0.03	112	3.31
Two or more races	23	0.04	15	0.03	0	0.00	8	0.24
Total	1,072	1.79	857	1.80	37	0.41	178	5.27
Self Reported Ethnicity								
Latino	48	0.08	23	0.05	14	0.16	11	0.33
Non-Latino	174	0.29	132	0.28	23	0.26	19	0.56
Total	222	0.37	155	0.33	37	0.41	30	0.89
Completed interviews	59,938	100.00	47,614	100.00	8,945	1.00	3,379	100.00

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

8.4.2 Imputation of the OMB Race-Ethnicity Variable

The DOF control totals are defined in terms of OMB race categories for raking dimensions 5, 6, and 8. Persons who reported themselves as Latino “some other race” were assigned an OMB race following procedures similar to those used in Census 2000. Since the OMB assignment is done using the imputed regular single race variables, all sampled persons have nonmissing races values for variables SRW, SRAA, SRAI, SRAS, SRPI, and SRO.

The OMB race-ethnicity variable, OMBSRREO, was assigned as follows:

- If the person self-reported as Latino (SRH=1), the variable OMBSRREO was set to 1. This assignment is independent of the values of the race variables.

- If the person self-reported as non-Latino (SRH=2) and reported OMB race alone or in combination with one or more OMB races (e.g., White alone, White and Black or African American, White and Black or African American and American Indian and Alaska Native) then OMBSRREO was given the value 2, 3, 4, 5, 6 or 7 (see Table 8-5) depending on the values of SRW, SRAA, SRAI, SRAS, and SRPI. In other words, there is no modification of race for non-Latinos who reported a valid OMB race(s).

- If the person self-reported as non-Latino (SRH=2) and reported both an OMB race and “some other race” (SRO=1), then OMBSRREO was assigned using only the specified OMB race(s). For example, non-Latino White and some other race became non-Latino White alone. This scenario is an example of the differences between OMBSRREO and the regular race-ethnicity variables (SRH, SRW, SRAA, SRAI, SRAS, and SRPI). Persons who reported two races, with one of them “some other race” are considered as single race respondents based on the OMB definition.

After the race/ethnicity assignments were made, 98 persons (0.16 percent) remained with missing values of OMBSRREO. These persons self-reported as non-Latino and other race only (SRH=2 and SRO=1). The missing values were imputed using the same procedures used to impute the regular single race variables as described above. In this case, temporary OMB race variables named SRW2, SRAA2, SRAI2, SRAS2, and SRPI2 were created using the values of already imputed SRW, SRAA, SRAI, SRAS, and SRPI. The values of the temporary OMB race variables were set to missing for the cases where the person self-reported as non-Latino and other race only. The missing values were imputed through a series of hot-deck imputations where pools of donors were created by matching the structure of the household and non-missing values of race and ethnicity of the adult, child, or adolescent in the household within geographic areas (i.e., stratum, region, or urban/rural area). For cases where there was no pool of donors based on household structure, missing values were imputed using the value of SRW2, SRAA2, SRAI2, SRAS2, and SRPI2 from another member of the household. Next, the variable OMBSRREO was assigned for the records with SRH=2 and SRO=1 using the imputed of values SRW2, SRAA2, SRAI2, SRAS2, and SRPI2. Table 8-8 shows the counts and percentages of imputed OMBSRREO values by type of extended interview (adult, child, and adolescent).

Table 8-8. Number and percentage of completed interviews with missing OMB race and ethnicity by extended interview type

OMB Race-ethnicity (OMBSRREO)	Total		Extended interview type: Adult		Child		Adolescent	
	Imputed Count	%	Imputed Count	%	Imputed Count	%	Imputed Count	%
1. Latino	NA		NA		NA		NA	
2. Non-Latino White alone	77	0.13	69	0.14	7	0.08	1	0.03
3. Non-Latino African American alone	7	0.01	5	0.01	1	0.01	1	0.03
4. Non-Latino Asian alone	1	0.00	1	0.00	0	0.00	0	0.00
5. Non-Latino American Indian/ Alaska Native alone	7	0.01	6	0.01	1	0.01	0	0.00
6. Non-Latino Native Hawaiian and Other Pacific Islander alone	0	0.00	0	0.00	0	0.00	0	0.00
7. Non-Latino two or more races	6	0.01	5	0.01	1	0.01	0	0.00
Total	98	0.16	86	0.18	10	0.11	2	0.06
Completed interviews	59,938		47,614		8,945		3,379	

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

8.4.3 Self-Reported Asian Ethnic Group

The person weights were raked using a dimension defined for Asian groups (Dimension 7). Since there was only one weight for the combined landline and supplemental list samples in CHIS 2009, we added a variable (OMBSRASO) for a raking dimension that would improve the estimates of the largest Asian ethnic groups in California. The variable OMBSRASO identifies the OMB non-Latino Asian ethnic group and is defined in Table 8-9.

Table 8-9. OMB Non-Latino Asian ethnic groups (OMBSRASO)

OMBSRASO	Description
1	Non-Latino Chinese alone
2	Non-Latino Korean alone
3	Non-Latino Filipino alone
4	Non-Latino Vietnamese alone
5	Other

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

The process to derive the variable OMBSRASO used the temporary OMB race variable SRAS2 previously created for the imputation of OMBSRREO. For records where SRAS2=1 (self-reported as OMB Asian alone or combined with some other race), five flags indicating the Asian ethnic groups of the respondent were derived using the Asian ethnic group questions in the extended interview (questions AA5E_1 to AA5E_18 for adults, TI7_1 to TI 7_18 for adolescents, and CH7_1 to CH7_18 for children). The name and description of the Asian ethnic group flags are shown in Table 8-10.

Table 8-10. OMB Asian group flags

Variable	Description
SRCH	Self-reported Chinese
SRPH	Self-reported Filipino
SRKR	Self-reported Korean
SRVT	Self-reported Vietnamese
SRASO	Self-reported Other Asian ethnic group

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

The missing values of the OMB Asian group variables (SRCH, SRPH, SRKR, SRVT, and SRASO) were imputed in the same way as the OMB race variables (or the temporary OMB race variables). A series of hot-deck imputations were run where pools of donors were created by matching the structure of the household and non-missing values of race, ethnicity, and Asian ethnic group of the adult, child, or adolescent in the household within geographic areas (i.e., stratum, region, or urban/rural area). For cases where there was no pool of donors based on household structure, race, ethnicity and Asian ethnic group, missing values were imputed using the values of SRCH, SRPH, SRKR, SRVT, and SRASO from another member of the household. The variable OMBSRASO was then created using the variables SRH, SRAA2, SRAI2, SRAS2, SRPI2, and the variables SRCH, SRPH, SRKR, SRVT, and SRASO after imputation. Table 8-11 shows the counts and percentages of imputed OMBSRASO values by type of extended interview (adult, child, and adolescent).

Table 8-11. Number and percentage of completed interviews with imputed OMB Asian ethnic group by extended interview type

			Extended interview type					
	Total		Adult		Child		Adolescent	
OMB Asian group (OMBSRASO)	Imputed count	%	Imputed count	%	Imputed count	%	Imputed count	%
1. Non-Latino Chinese	1	0.00	0	0.00	1	0.01	0	0.00
2. Non-Latino Korean	1	0.00	1	0.00	0	0.00	0	0.00
3. Non-Latino Filipino	1	0.00	1	0.00	0	0.00	0	0.00
4. Non-Latino Vietnamese	1	0.00	1	0.00	0	0.00	0	0.00
5. Other	94	0.16	83	0.17	9	0.10	2	0.06
Total	98	0.16	86	0.18	10	0.11	2	0.06
Completed interviews	59,938		47,614		8,945		3,379	

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

8.5 Self-Reported County and Self-Reported Stratum

In CHIS 2009, the geographic location variables such as self-reported county of residence, Los Angeles SPA, San Diego County Health Region, Census tract, and self-reported stratum were assigned after geocoding the geographic information collected during the interview (address of respondent, nearest street intersection, or self-reported county) or attached to the sample telephone number (the mailing address or ZIP Code covered by the telephone exchange). Table 8-12 shows the variables used in the geocoding process.

Table 8-12. Variables used in geocoding

Variable	Description	Source
AH42	County of residence (self report)	Adult questionnaire
AO1ADDR	Confirmed/corrected street address	Adult questionnaire
AO1CITY	Confirmed/corrected city	Adult questionnaire
AO1ZIP	Confirmed/corrected ZIP Code	Adult questionnaire
AM7	ZIP Code (self report)	Adult questionnaire
AO2ADDR	Street address (self report)	Adult questionnaire
AO2CITY	City (self report)	Adult questionnaire
AM8	Street name of residence (self report)	Adult questionnaire
AM9	Street name of nearest cross street (self report)	Adult questionnaire
M_ADDR	Street address (matched to phone number prior to interview)	Address mailing vendor
M_CITY	City (matched to phone number prior to interview)	Address mailing vendor
M_ZIP	ZIP Code (matched to phone number prior to interview)	Address mailing vendor
S_ZIP	ZIP Code (provided by sample vendor for every phone)	Sample vendor

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

The derived location variables SRSTRATA (self-reported stratum), SRCOUNTY (self-reported county), SR_LASPA (self-reported Los Angeles SPA), SR_HR (self-reported San Diego County Health Region) are household-level variables that were assigned to all adult, child and adolescent records within the same household before creating the raking dimensions. The variable SRSTRATA was used to create the cells for raking dimensions 1, 2, and 8 defined at the stratum or California region level while the variables SRCOUNTY (self-reported county), SR_LASPA (self-reported Los Angeles SPA), and SR_HR (self-reported San Diego County Health Region) were used to create the cells for raking dimension 4 defined for Los Angeles County and San Diego County.

Table 8-13 shows the distribution of adult respondents by self-reported stratum compared with the sampling stratum for the landline/surname sample. Each stratum had migration in and migration out as a result of self-reports not matching the sampling stratum. This table shows that the net effect of cross-stratum migration is small, with the greatest differences for strata with the lowest geographic counts, as indicated by the net agreement ratios (NAR) in the rightmost column of Table 8-13. The NAR is the number of respondents in the sampling stratum divided by the number of respondents in the self-reported stratum. A NAR value less than one indicates more in-migration than out-migration from the stratum, and a value greater than one the reverse. Most values are very close to one, indicating either very little migration or roughly equivalent rates of in- and out-migration.

Table 8-13. Distribution of self-reported strata and sampling strata for the landline/surname samples

Stratum	Sampling stratum	Self-reported stratum	Net agreement ratio
Los Angeles	11,097	11,112	1.00
San Diego	6,297	6,302	1.00
Orange	3,361	3,301	1.02
Santa Clara	2,155	2,219	0.97
San Bernardino	1,884	1,883	1.00
Riverside	2,040	2,069	0.99
Alameda	1,557	1,507	1.03
Sacramento	1,529	1,529	1.00
Contra Costa	1,104	1,174	0.94
Fresno	877	870	1.01
San Francisco	882	871	1.01
Ventura	1,136	1,166	0.97
San Mateo	739	696	1.06
Kern	768	762	1.01
San Joaquin	676	674	1.00
Sonoma	653	666	0.98
Stanislaus	616	586	1.05
Santa Barbara	755	741	1.02
Solano	596	598	1.00
Tulare	617	629	0.98
Santa Cruz	646	618	1.05
Marin	2,521	2,504	1.01
San Luis Obispo	565	571	0.99
Placer	638	642	0.99
Merced	651	679	0.96
Butte	590	613	0.96

Table 8-13. Distribution of self-reported strata and sampling strata for the landline/surname samples (continued)

Stratum	Sampling stratum	Self-reported stratum	Net agreement ratio
Shasta	600	630	0.95
Yolo	700	682	1.03
El Dorado	602	625	0.96
Imperial	728	720	1.01
Napa	575	585	0.98
Kings	641	639	1.00
Madera	718	713	1.01
Monterey	553	637	0.87
Humboldt	1,005	1,009	1.00
Nevada	640	626	1.02
Mendocino	717	693	1.03
Sutter	581	593	0.98
Yuba	608	556	1.09
Lake	644	638	1.01
San Benito	745	688	1.08
Colusa, Glenn, Tehama	483	451	1.07
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	474	496	0.96
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	454	455	1.00

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

Table 8-14 shows the distribution of adult respondents by self-reported stratum compared with the sampling stratum for the cell phone sample (including the ported numbers). In this case, the net effect of cross-stratum migration is larger, with the greatest differences for Sacramento and Northern and Sierra areas. However, during sample design these migrations were considered when drawing the sample. As a result, the cases by self-reported region are closer to the target goals by region.

Table 8-14. Distribution of self-reported strata and sampling strata for the cell phone sample

Stratum	Sampling stratum	Self-reported stratum	Net agreement ratio
State	3,728	3,728	1.00
Northern & Sierra Counties	812	657	1.24
Greater Bay Area	599	567	1.06
Sacramento Area	258	472	0.55
San Joaquin Valley	668	627	1.07
Central Coast	486	469	1.04
Los Angeles	491	493	1.00
Other Southern California	414	443	0.93

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

9. VARIANCE ESTIMATION

This chapter describes the methods and results of computing sampling errors for CHIS 2009 data. The first section gives an overview of the reason for computing sampling errors and summarizes the precision of estimates for adults, children, and adolescents produced from the weights that include the landline, list, and cell phone samples. The remainder of the chapter describes the methodology for producing estimates of sampling variability. Section 9.2 is a general review of the two main methods of computing sampling errors or variances of estimates from surveys with complex sample designs like CHIS 2009. Section 9.3 describes a replication method of variance estimation that can be used with the data. Section 9.4 shows how analysts can compute sampling errors for CHIS 2009 estimates using commercially available software.

9.1 Design Effects

To evaluate the precision of sample estimates derived from a survey, sampling errors are computed. Estimates of sampling errors can be used to make inferences about the size of the difference between two population parameters based on the values of corresponding sample estimates, their estimated precision, and the expected probability distribution of such a difference. Suppose an analyst wishes to compare the proportion of employed persons whose employer offers health care benefits in two counties in California. By taking the estimated sampling error of this difference into account, the analyst can make inferences about the size of the difference.

Inferences of this nature require an estimate of the precision or sampling error of the characteristic being investigated. There is a variety of ways of reporting the estimated precision of a survey estimate including:

- A standard error (the standard deviation of the estimate);
- A variance of an estimate (the standard error squared);
- A coefficient of variation (the ratio of the standard error to the estimate); or
- A confidence interval (the estimate plus or minus a multiple of the standard error).

Another way of describing the variability of an estimate from a survey is by using the "design effect." The concept of a design effect was introduced and popularized by Kish (1965) to account for the additional variability associated with complex sample designs involving stratification and clustering. The design effect is the ratio of the variance of the sample estimate for the survey (with its particular sample design and estimation method) to the variance of a simple random sample of the same sample size.

For a specific sample, the design effect, or *DEFF*, for an estimate from a survey can be estimated as

$$DEFF = \frac{\text{sampling variance of a complex sample}}{\text{sampling variance of a simple random sample}}.$$

At the analysis stage, the *DEFF* is useful because many procedures in statistical software assume the data are from a simple random sample when computing sampling errors of estimates. The *DEFF* can, in some circumstances, indicate the appropriateness of this assumption and can be used to adjust the sampling errors of the estimates to produce ones that are closer to the actual sampling errors (Skinner, Holt, and Smith, 1989).

Calculating the design effect for a proportion is straightforward because the variance of an estimated proportion in a simple random sample can be estimated easily. In this case, the estimated *DEFF* for a proportion is

$$DEFF_{PROP} = \frac{v(\hat{p})_{COMPLEX}}{v(\hat{p})_{SRS}},$$

where $\hat{p}$ is the estimated proportion, $v(\hat{p})_{SRS}$ is the variance estimate of the estimated proportion assuming a simple random sample, and $v(\hat{p})_{COMPLEX}$ is the variance of the estimated proportion accounting for the complex sample survey design.

In most surveys, design effects are larger than one. In CHIS 2009, design effects are greater than one mainly because the cases have different estimation weights (Kish, 1992). As will be seen shortly, design effects from the survey are considerably greater than one for some statewide estimates.

Design effects are of primary interest to data users. They reveal that the complex sample design and estimation procedures used resulted in estimates of variances that are greater than what would be obtained from a simple random sample. A simple random sample design was not considered for CHIS 2009, because it would not have achieved the sample sizes for the specific domains of interest, in particular at the county/stratum level, for given resources. The design effects calculated from the CHIS 2009 data indicate that the sample design used in the survey needs to be taken into account when analyzing the data.

In CHIS 2009, as in most large-scale surveys, a large number of data items are collected. Each resulting variable has its own design effect. One way to summarize the design effects for the items is to compute *DEFF*s for a number of items and then average them. This average represents the design effects for similar items from the survey, as described in Wolter (1985).

The *DEFT* is the square root of the design effect, and it is similar to the *DEFF* but on the scale of the standard error of the estimate rather than the variance. Taking the square root of the *DEFF* has a smoothing affect on the variability.

The tables in the following sections show the *DEFF*s and *DEFT*s for selected items from either the adult, the child or the adolescent interviews. The *DEFT* is often considered a more convenient measure than the *DEFF*, because it can be used directly when computing confidence intervals for the estimates. See Verma, Scott, and O'Muircheartaigh (1980) for a discussion of the use of the *DEFT*. The main reason for presenting the *DEFT*s here is because it dampens some of the noise associated with the *DEFF*s. The maximum and minimum values of the *DEFF*s in the tables show that there is considerable variability in these quantities.

Before reviewing the tables in detail, it is important to discuss the most important factors that result in design effects larger than one. These factors are

- **Oversampling.** For the landline/list sample, the need for both county and state estimates required oversampling to produce stable estimates for these areas. This oversampling increased the design effect for statewide estimates. The cell sample had minor disproportionate sampling. However, when the samples are combined, persons in cell only households were subsampled.

- **Within-Household Subsampling.** For all samples only one adult was selected in each household. In the landline/list sample one child and/or adolescent was sampled in each household. This subsampling contributed to the differential weights at the person level because persons in households with more persons were subsampled at lower rates. In addition, young children (age 0 to 5 years) were sampled at twice the rate of older children (age 6 to 11 years)

- **Weighting Adjustments.** Differential weights were applied to reduce nonresponse bias and to make the estimates consistent with known population totals. The main reason for including these adjustments was to reduce biases in the estimates, but some of the adjustments may have increased the design effects for some estimates.

- **Composite weight.** The CHIS 2009 weights combine samples from overlapping domains that were sampled at different rates.

9.1.1 Design effect for the combined sample weights

Tables 9-1 to 9-3 present the *DEFF*s and *DEFT*s of the adult, child and adolescent interviews, respectively, for the landline/list/cell sample. The first panel in the tables shows the average, median, minimum, and maximum *DEFF*s computed for a combination of categorical and continuous variables. The rightmost panel shows the average *DEFT* for the same items. The *DEFF*s and *DEFT*s were calculated using 15 items selected from the adult interview, 11 items from the child interview, and 17 items from the adolescent interview. The variables include health characteristics such as general health rating, diagnosis (i.e, asthma, diabetes, high blod pressure, heart failure/congestive, heart disease, difficulty learning and remembering, child have teeth, child visited emergency room, felt nerveous, had psychological or emotional counseling), live style (exercise, smoking and alcohol, go to the park, had fast food), preventive medicine (mammogram, blood test, flu vaccine, delayed medical care, child care), health insurance (insured, employer health insurance, other government helath plan, prescription coverage), and socio economic and demographic variables (skipped meals, income, sexual orientation, marital status, education attainment, employed, servings of juice and vegetables, attended school last week). All were calculated by stratum.

Table 9-1 shows that in 33 counties the average *DEFT*s for estimates of adult items are between 1.17 and 1.7 . This implies that for 75 percent of the strata the standard error of the estimates is about 17 to 70 percent greater than the expected standard error of a simple random sample. The average *DEFT* for the state estimates is 2.05. This is larger than the county-level *DEFT*s as expected (except for San Benito) because counties were not sampled proportional to their population.

Table 9-1. Average *DEFF* and *DEFT* for estimates from the adult interview

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
State	4.32	4.31	6.87	1.45	2.05
Los Angeles	4.18	3.62	7.81	1.35	2.01
San Diego	4.50	3.72	15.42	0.85	2.02
Orange	3.85	2.99	8.16	1.81	1.90
Santa Clara	3.35	3.20	8.21	1.10	1.78
San Bernardino	3.19	3.49	4.59	0.68	1.75
Riverside	3.72	3.91	6.73	0.32	1.85
Alameda	3.27	3.02	5.89	1.21	1.75
Sacramento	2.45	2.15	4.78	0.81	1.52
Contra Costa	2.50	2.44	4.28	0.45	1.54
Fresno	1.96	1.72	4.67	0.40	1.35
San Francisco	2.92	3.04	4.36	0.68	1.68
Ventura	2.05	2.05	3.51	0.16	1.39
San Mateo	2.13	1.76	4.24	0.61	1.41
Kern	3.27	2.72	8.26	0.40	1.72
San Joaquin	1.83	1.88	3.16	0.69	1.32
Sonoma	1.95	1.94	2.91	0.61	1.37
Stanislaus	1.60	1.69	2.28	0.30	1.24
Santa Barbara	1.97	1.66	3.45	0.23	1.36
Solano	2.64	2.30	6.33	1.21	1.57
Tulare	1.88	1.88	2.77	0.69	1.35
Santa Cruz	1.81	1.87	2.52	0.21	1.32
Marin	4.53	4.21	10.33	0.23	1.98
San Luis Obispo	2.18	1.95	5.02	0.27	1.41
Placer	1.51	1.63	2.21	0.33	1.21
Merced	2.23	2.27	3.80	0.25	1.45
Butte	1.87	1.40	7.88	0.34	1.28
Shasta	3.00	2.50	11.00	0.62	1.63
Yolo	2.46	2.10	5.68	0.69	1.52
EL Dorado	1.71	1.65	3.10	0.47	1.27
Imperial	1.61	1.69	2.82	0.60	1.25
Napa	2.73	2.57	4.28	0.16	1.60
Kings	3.50	2.90	12.48	0.42	1.75
Madera	1.77	1.88	2.66	0.25	1.30
Monterey	1.71	1.59	2.43	0.86	1.29
Humboldt	3.04	2.47	8.33	1.33	1.69
Nevada	1.75	1.52	3.97	0.28	1.28
Mendocino	1.73	1.91	2.59	0.34	1.28
Sutter	1.42	1.49	2.21	0.48	1.17
Yuba	2.43	2.35	5.06	0.19	1.49
Lake	1.59	1.63	2.12	0.67	1.25
San Benito	4.77	3.90	16.61	1.75	2.08
Colusa, Glen, Tehama	1.51	1.45	2.89	0.29	1.19
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	1.71	1.72	3.08	0.26	1.28
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.68	1.55	3.98	0.51	1.26

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 9-2 shows the average *DEFT* for estimates from the child interview in each stratum for the landline/list weights. The average *DEFT* at the state level is 1.70. In approximately 84 percent of the counties, the average *DEFT*s for the counties vary between 1.12 and 1.50 ; that is, the standard errors of these estimates are between 10 and 50 percent greater than expected from a simple random sample. Unlike previous cycles of CHIS, the state average *DEFT*s for estimates from the child interview are smaller than those for the adult interview. This result was not expected; because the subsampling at the person level for children is typically more variable than it is for adults (the number of children per household is more variable than the number of adults per household).

Table 9-2. Average *DEFF* and *DEFT* for estimates from the child interview

	Design effect (*DEFF*)				*DEFT*
Stratum	Average	Median	Maximum	Minimum	Average
State	2.96	3.02	4.28	1.65	1.70
Los Angeles	3.15	3.29	5.11	1.40	1.73
San Diego	1.99	1.90	2.73	1.44	1.40
Orange	3.93	4.42	6.76	0.75	1.92
Santa Clara	2.49	2.43	3.22	2.10	1.57
San Bernardino	2.02	2.05	2.45	1.53	1.42
Riverside	2.22	1.86	4.10	1.07	1.45
Alameda	1.85	1.80	2.54	0.99	1.35
Sacramento	1.80	1.86	2.68	0.20	1.30
Contra Costa	1.91	2.00	2.65	0.75	1.36
Fresno	1.68	1.73	2.63	0.33	1.27
San Francisco	2.13	2.28	2.85	1.12	1.45
Ventura	2.15	2.16	2.96	0.78	1.44
San Mateo	1.38	1.56	1.79	0.68	1.16
Kern	2.12	2.07	3.71	0.97	1.43
San Joaquin	2.16	2.19	3.23	0.73	1.44
Sonoma	1.62	1.86	2.01	0.23	1.24
Stanislaus	1.86	1.64	3.98	1.11	1.34
Santa Barbara	1.51	1.52	1.98	0.73	1.22
Solano	1.54	1.45	2.32	0.36	1.21
Tulare	1.34	1.37	1.95	0.36	1.13
Santa Cruz	2.17	2.23	3.28	1.11	1.46
Marin	1.48	1.52	1.92	0.86	1.21
San Luis Obispo	2.10	2.14	4.60	0.67	1.40
Placer	1.73	1.44	3.09	0.94	1.29
Merced	1.43	1.47	1.97	0.92	1.19
Butte	1.62	1.44	2.87	1.19	1.26
Shasta	1.36	1.54	1.91	0.82	1.16
Yolo	1.74	1.72	2.00	1.57	1.32
EL Dorado	1.41	1.38	2.37	0.76	1.17

Table 9-2. Average *DEFF* and *DEFT* for estimates from the child interview (continued)

Stratum	Design effect (*DEFF*) Average	Median	Maximum	Minimum	*DEFT* Average
Imperial	3.23	2.81	6.39	0.80	1.72
Napa	1.28	1.38	1.68	0.70	1.12
Kings	1.32	1.41	1.59	0.70	1.14
Madera	2.39	2.45	5.11	0.75	1.50
Monterey	1.32	1.27	1.71	0.97	1.14
Humboldt	1.40	1.28	1.90	0.89	1.17
Nevada	1.82	1.75	2.37	1.16	1.34
Mendocino	2.19	2.65	4.07	0.50	1.43
Sutter	2.28	2.20	2.85	2.07	1.51
Yuba	1.83	1.88	2.44	1.03	1.34
Lake	1.52	1.43	2.67	0.68	1.21
San Benito	2.44	1.93	4.41	1.13	1.53
Colusa, Glen, Tehama	2.29	2.53	4.11	0.65	1.47
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	1.92	2.38	3.49	0.27	1.30
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	2.92	2.41	7.48	1.01	1.65

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 9-3 shows that the average *DEFT* for items from the adolescent interviews are similar to those from the child interviews. Since the sampling for adolescents is similar to that for children, we expect a close correspondence between the two. The state average *DEFT*s are lower for adolescents than for children, primarily because there was no oversampling of adolescents by age and there are fewer adolescents than children per household, both of which reduce the variability in the weights. The average *DEFT* for the state estimates is 1.44. In 34 strata (77 percent) the average *DEFT*s are between 1.02 and 1.30.

Table 9-3. Average *DEFF* and *DEFT* for estimates from the adolescent interview

Stratum	Design effect (*DEFF*) Average	Median	Maximum	Minimum	*DEFT* Average
State	2.09	2.06	2.65	1.35	1.44
Los Angeles	1.91	1.90	2.49	1.32	1.38
San Diego	1.47	1.61	1.97	0.69	1.20
Orange	2.02	2.24	3.20	0.03	1.36
Santa Clara	1.85	1.89	3.14	0.65	1.34
San Bernardino	2.61	2.33	6.40	0.69	1.57
Riverside	1.23	1.35	1.77	0.50	1.09
Alameda	1.91	1.93	2.81	1.19	1.37
Sacramento	1.64	1.72	2.30	0.61	1.27

Table 9-3. Average *DEFF* and *DEFT* for estimates from the adolescent interview (continued)

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Contra Costa	2.48	2.00	5.57	1.08	1.51
Fresno	1.21	1.18	1.69	0.56	1.09
San Francisco	1.31	1.41	2.25	0.77	1.13
Ventura	1.47	1.42	2.15	0.77	1.19
San Mateo	1.65	1.67	2.78	0.58	1.27
Kern	1.63	1.44	3.53	1.00	1.25
San Joaquin	1.42	1.43	1.89	1.12	1.19
Sonoma	1.30	0.93	2.66	0.57	1.10
Stanislaus	1.14	1.13	1.60	0.81	1.06
Santa Barbara	1.49	1.40	2.89	0.62	1.20
Solano	1.63	1.79	2.05	1.12	1.27
Tulare	1.79	2.26	3.00	0.35	1.27
Santa Cruz	1.46	1.77	2.13	0.50	1.18
Marin	1.97	1.75	3.80	1.30	1.38
San Luis Obispo	0.94	1.01	1.14	0.44	0.97
Placer	1.25	1.24	1.79	0.68	1.11
Merced	1.48	1.31	2.20	0.85	1.20
Butte	1.27	1.29	1.86	0.93	1.12
Shasta	1.51	1.50	2.33	0.91	1.22
Yolo	2.37	2.52	3.74	1.14	1.52
EL Dorado	1.18	1.20	1.64	0.83	1.08
Imperial	1.79	1.28	3.27	0.76	1.29
Napa	1.29	1.29	1.62	0.98	1.13
Kings	1.31	1.45	1.70	0.57	1.13
Madera	1.60	1.65	1.83	1.16	1.26
Monterey	1.00	0.97	1.54	0.50	0.99
Humboldt	1.37	1.49	1.81	0.76	1.17
Nevada	1.08	0.98	1.56	0.72	1.03
Mendocino	1.06	1.03	1.44	0.46	1.02
Sutter	1.12	1.24	1.68	0.61	1.04
Yuba	1.69	1.84	2.33	0.49	1.27
Lake	1.23	1.20	1.70	0.60	1.10
San Benito	1.25	1.09	2.28	0.43	1.08
Colusa, Glen, Tehama	1.28	1.50	2.11	0.35	1.10
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	1.51	1.44	2.59	0.96	1.22
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.18	1.21	1.82	0.61	1.08

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

9.2 Methods for Variance Estimation

Variance estimation procedures have been developed to account for the complex sample design. Using these procedures, factors such stratification, multistage sampling, sampling from different frames, and the use of differential sampling rates to oversample a targeted subpopulation can be appropriately reflected in estimates of sampling error. The two main methods are replication and linearization or the Taylor series approximation. Wolter (1985) is a useful reference on the theory and applications of these methods. Shao (1996) is a more recent review paper that compares these methods. The rest of this section briefly reviews these methods.

The basic idea behind replication is to draw subsamples from the sample, compute the estimate from each of the subsamples, and estimate the variance of the original sample using the variability of the subsample estimates. Specifically, subsamples of the original "full" sample are selected to calculate subsample estimates of a parameter for which a "full-sample" estimate of interest has been generated. The variability of these subsample estimates about the estimate for the full sample can then be assessed. The subsamples are called replicates, and the estimates from the subsamples are called replicate estimates. Rust and Rao (1996) discuss balanced repeated replication (BRR) and jackknife replication, two general approaches to forming subsamples. They show how the units included in the subsample can be defined using variance strata and units. They also describe how these methods can be implemented using replicate weights.

Replicate weights are created to produce the corresponding replicate estimate. Each replicate weight is computed using the same estimation steps as the full sample weight but using only the subsample of cases comprising each replicate. Once the replicate weights are developed, it is a straightforward matter to compute variance estimates for sample estimates of interest. The variance estimate takes the following form:

$$v\left(\hat{\theta}\right) = c \sum_{k=1}^{G} \left(\hat{\theta}_{(k)} - \hat{\theta}\right)^2 \qquad (1)$$

where

- θ is an arbitrary parameter of interest;
- $\hat{\theta}$ is the estimate of θ based on the full sample;
- $\hat{\theta}_{(k)}$ is the k^{th} estimate of θ based on the observations included in the k^{th} replicate;
- G is the total number of replicates formed;
- c is a constant that depends on the replication method; and

$v\left(\hat{\theta}\right)$ is the estimated variance of θ.

The other widely used method for variance estimation for complex sample surveys is called linearization and is based on the Taylor series approximation. In this method, the Taylor series linearization of a statistic is formed and then substituted into the formula for calculating the variance of a linear estimate appropriate for the sample design. Linearization relies on the simplicity associated with estimating the variance for a linear statistic even with a complex sample design.

9.3 Design of Replicates

In CHIS 2009, a paired unit jackknife method (JK2[15]), a form of jackknife replication, was selected for computing variances. This section provides details on setting up the replication structure, including the definition of the variance strata and units.

Two major reasons for using replication to estimate variances for CHIS 2009 are operational convenience and the ability to reflect all components of the design and estimation in the estimates of variability. With respect to operational convenience, once replicate weights are constructed, it is very simple to compute estimates of sampling errors. No special care is needed for subgroups of interest, and no knowledge of the sample design is required. If an estimator is needed that was not previously considered, replication methods can be easily used to develop an appropriate estimate of variance. In such a case, variance estimates using a Taylor series approach would require additional work. The variance estimation stratum and unit must also be included in the file for the Taylor series method.

The second reason for using replication is probably more important. The nonresponse, composite factor, and raking types of adjustments made in developing the CHIS 2009 analysis weights all affect the sampling errors of the estimates produced from the survey.

Furthermore, the set of weights created in CHIS 2009 combined samples from different frames and were raked to the same control totals. The replicate weights prepared for CHIS reflect all such aspects of weighting and raking. Currently existing software for using the Taylor

[15]This method is denoted as JK2 in the software program, WesVar, which was used to compute all the sampling errors in this report.

series method for variance estimation cannot reflect these weighting adjustments. In some Taylor series software poststratification can be taken into account, but only in specific situations.

In the JK2 replication method, adjacent pairs of sampled telephone numbers are treated as having been sampled from the same stratum. Each pair of sampled telephone numbers is treated as an implicit stratum, where each such stratum is defined by the sort order used in the sample selection of telephone numbers. In this method, the constant, c, in equation (1) equals 1. This approach has been used in previous cycles of CHIS and in other RDD studies such as the National Household Education Survey, an ongoing national RDD survey starting in 1991 and with a most recent cycle in 2007 (Hagedorn, et. al., 2008).

The first step in designing the replicate structure is to determine the number of variance estimation strata. In the JK2 method, the number of replicates is equal to the number of variance estimation strata. The choice of the number of variance estimation strata is based on the desire to obtain an adequate number of degrees of freedom to ensure stable estimates of variance while not having so many as to make the cost of computing variance estimates unnecessarily high. Generally, at least 30 degrees of freedom are needed to obtain relatively stable variance estimates. A number greater than 30 is often targeted because there are other factors that reduce the contribution of a replicate to the total number of degrees of freedom, especially for estimates of subgroups.

For CHIS 2009 and previous cycles of CHIS, we elected to create 80 variance estimation strata, even though many more could have been created. For the landline and cell phone samples, the 80 variance strata were formed as follows. First, the sampled telephone numbers were arranged in the same sort order used in sample selection. Next, adjacent sampled telephone numbers were paired to establish initial variance estimation strata (the first two sampled phone numbers were the first initial stratum, the third and fourth sampled telephone numbers were the second initial stratum, etc). Each telephone number in the pair was randomly assigned to be either the first or second variance unit within the variance stratum. Each pair was sequentially assigned to one of 80 final variance estimation strata (the first pair to variance estimation stratum 1, the second to stratum 2, ..., the 80th stratum pair to stratum 80, the 81st pair to stratum 1, etc.). As a result, each variance stratum had approximately the same number of telephone numbers. The same process was followed for each sampling stratum.

Once the variance strata are created, the replicate weights can be created. The full replicate weights are constructed by first modifying the full sample base weights. The replicate base weight for replicate *k* for record *i* is

$$w_i^{(k)} = \begin{cases} 2w_i, \text{ if } i \text{ is in variance stratum } k \text{ and variance unit 1} \\ 0, \text{ if } i \text{ is in variance stratum } k \text{ and variance unit 2} \\ w_i, \text{ if } i \text{ is not in variance stratum } k \end{cases}$$

The same sequence of weighting adjustments used in the full sample weight is then applied to the replicate base weights to create the final replicate weights. Thus, all of the different components of the weighting process are fully reflected in the replicate weights, ranging from household adjustments (nonresponse, adjustment for household noncoverage, and adjustment to control totals) to person adjustments (nonresponse and raking).

9.4 Software for Computing Variances

In the past, most standard statistical software packages assumed a simple random sample when computing estimates of variance. As a result, estimates of variance from these packages had the potential to seriously understate the true variability of the survey estimates. However, in recent years, specialized commercial software has been developed to analyze data from complex surveys (Lepkowski and Bowles, 1996). In this section, we describe the elements needed to compute estimates for CHIS 2009 using some of these programs.

WesVar Version 4.3 (Westat, 2000) is a free software package developed and distributed by Westat. WesVar uses replication methods to compute variance estimates. WesVar is an interactive program with a graphical interface that makes it simple to specify the estimates for sampling errors for estimates of interest. The data requests center on sessions called "workbooks." A workbook is a file linked to a specific WesVar data set. In a workbook, the user can request descriptive statistics, as well as analyze and create new statistics. Descriptive statistics of analysis variables are produced through "table requests." Regression requests support both linear and logistic regression models. Outputs include statistics of interest, such as the sum of weights, means, percentages, along with their corresponding standard errors, design effects, coefficients of variation, and confidence intervals.

To use WesVar with CHIS 2009 data, the only requirements are to identify the full and replicate weights that are on the data file and specify the replication method as JK2. This specification is made when a workbook is opened. All of the standard errors produced will properly account for the sample design and estimation methods because these features are accounted for in the replicate weights.

SUDAAN® (Research Triangle Institute, 2008) is a package developed by Research Triangle Institute to analyze data from complex sample surveys. SUDAAN is available as a standalone package or it can be called using SAS. SUDAAN and WesVar produce the same point estimates. The difference between the two packages is in the method used to compute the variances. While WesVar uses replication exclusively, SUDAAN can use either a first-order Taylor series expansion approximation (linearization), or replication. When the Taylor series approximations are used, SUDAAN does not fully take into account complex weighting schemes such as nonresponse adjustments or raking, so the variance estimates will be different than estimates calculated using replication. On the other hand, if the user specifies replication as the variance estimation method, the estimates of variance computed in SUDAAN will take into account the sample design and weighting.

For descriptive statistics, SUDAAN offers two procedures: PROC CROSSTAB for categorical variables and PROC DESCRIPT for continuous variables. These procedures can be used to compute statistics of interest, such as sum of weights, means, and percentages along with their corresponding standard errors, design effects, and confidence intervals. Both procedures use the option DESIGN= to specify the type of survey design when calculating variance estimates. If no design type is specified using this option, then a standard “with replacement” design is assumed and linearization is used for variance estimation. Specifying JACKKNIFE assumes the use of replication. In this instance, the WEIGHT and NEST statements are also required. SUDAAN also contains procedures for computing other analytic statistics, such as those associated with linear and regression models. Consult the help manuals (available online) for more detail on the procedures and options available for SUDAAN.

Beginning in Version 9.1, SAS® has also included procedures to analyze survey data. Version 9.2 (SAS Institute, 2004. In Version 9.3, these procedures can use either the linearization or replication methods (include the REPWEIGHTS statement) to estimate the variance. The procedures in SAS for analyzing survey data are SURVEYMEANS, SURVEYREG, SURVEYFREQ and SURVEYLOGISTIC. The SURVEYMEANS procedure computes estimates of means, proportions, percentiles, and totals, Estimates of differences or other linear combinations are not available in SURVEYMEANS. The SURVEYFREQ procedure produces one-way tand cross tabulation tables for survey data. This procedure also computes estimates of odds ratios and relative risk estimates. The SURVEYREG procedure fits linear regression models while SURVEYLOGISTIC performs logistic regression for survey data and fit various links including the cumulative logit, generalized logit, probit, and complementary log-log functions.

Another software package that can be used to analyze survey data is STATA 11 (StataCorp, 2009). STATA is a command driven, fully programmable statistical package used for managing, analyzing, and graphing data. STATA was developed by StataCorp and is available for a variety of platforms, including DOS, Windows, Macintosh, and UNIX. STATA's statistical, graphical, and data management capabilities are fully expandable through programming.

STATA has a family of *svy*- commands to analyze data from sample surveys. The set of analytic methods in STATA is more exhaustive than any other package. The *svy* commands can be used to estimate a variety of quantities such as totals, proportions, means, linear combinations of means, and logistic regression parameters. Two-dimensional tables of totals and proportions, along with *DEFF*s for proportions can be produced using *svy* **tab.** The command *svy* **mean** can be used to produce the *DEFF*s for proportions by coding the analytical variable with values 0 and 1. To estimate totals using *svy* **total**, a variable with a value of 1 must be created for all records in the file. The *svy* command in the latest version of STATA can perform general linear modeling (**glm** command), nonlinear least squares estimation (**nl** command), and conditional logistic regression (**clogit** command) among others.

Like SUDAAN and SAS, STATA can use linearization (**linear** variance type option) or replication (**jack** variance type option) to estimate variances. Besides point estimates (proportions, means, ratios and totals) and their standard errors, STATA can compute confidence intervals, design effects, and misspecification effects. Design and misspecification effects are computed for means and proportions only.

When using linearization to estimate variances the software packages referred to above require auxiliary variables that provide information about the sample design. Two variables have been defined and included in the data files (TSVARSTR and TSVRUNIT). TSVARSTR is required for all analyses, but TSVRUNIT is required only when analyses are performed using the combined data. In other words, when separate analyses are done by adults, children or teens the variable TSVRUNIT is not required. The definitions of TSVARSTR and TSVRUNIT are

- **TSVARSTR** (Taylor's series variance stratum). The variable TSVARSTR indicates the variance strata to be used for software that computes estimates of variance using the Taylor series method. In previous cycles of CHIS, the variable TSVARSTR was created by sequentially numbering the sampling strata. In CHIS 2009 where the landline, list and cell samples were combined, TSVARSTR was computed using the replicate variance stratum with values 1 to 80 from the landline/list and cell phone samples. While we believe this is a reasonable approach for linearization variances for combined samples with the type of adjustments made during weighting, linearization programs do not reflect the full range of weighting steps used in the production of the composite estimates.

- **TSVRUNIT** (Taylor's series unit). The variable TSVRUNIT indicates the PSU this case is the sampled household. In CHIS 2009, the value of TSVARUNIT corresponds to the variance unit created for the replicate weights.

The same variables, TSVARSTR and TSVRUNIT, can be used for linearization variance estimation in SUDAAN, SAS, and STATA.

REFERENCES

Blumberg SJ, Luke JV.(2008). Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2008. National Center for Health Statistics. Available from: http://www.cdc.gov/nchs/nhis.htm. December 17, 2008.

Blumberg, S.J., Luke, J.V., Cynamon, M. L. (2006). Telephone coverage and health survey estimates: Evaluating the need for concern about wireless substitution. *American Journal of Public Health*, 96(5), 926-931.

Brackstone, G.J., and Rao, J.N.K. (1979). An investigation of raking ratio estimation. *Sankhya,* 41, 97-114.

Brick, J. M., Flores Cervantes I., Lee S, and Norman G. (2010)**Error! Bookmark not defined.** Nonsampling errors in dual frame telephone surveys, *Survey Methodology*, Statistics Canada, June 2011 Vol. 37, No. 1, pp. 1-12

Brick, J.M., Judkins, D., Montaquila, J., and Morganstein, D. (2002). Two-phase list-assisted RDD sampling. *Journal of Official Statistics*, 18(2), 203-215.

Brick, J.M., and Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5, 215-238.

Brick, J. M., Montaquila, J., Hagedorn, M. C., Roth, S. B., and Chapman, C. (2005). Implications for RDD design from an incentive experiment, *Journal of Official Statistics*, 21(4), 571-589.

Council of American Survey Research Organizations. (1982). *Special report: On the definition of response rates*. Port Jefferson, NY: CASRO.

Deville, J.C., and Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association,* 87, 376-382.

Elliott, M. R., Little, R.J.A., and Lewitzky, S. (2000). Subsampling callbacks to improve survey efficiency. *Journal of the American Statistical Association,* 95(451), 730-738.

Ford, B. M. (1983). *An overview of hot-deck procedures*: *Vol. 2. Incomplete data in sample surveys*. New York: Academic Press.

Ford, E.S. (1998). Characteristics of survey participants with and without a telephone: Findings from the third national health and nutrition examination survey. *Journal of Clinical Epidemiology*, 51, 55-60.

Groves, R.M. (1989). *Survey errors and survey costs*. New York: John Wiley and Sons.

Hagedorn, M., Roth, S.B., O'Donnell, K. Smith, S., and Mulligan, G. (2008). National Household Education Surveys Program of 2007: *Data File User's Manual, Volume I.* (NCES 2009-024). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hansen, M.H., and Hurwitz, W.N. (1946). The problem of nonresponse in sample surveys. *Journal of the American Statistical Association*, 41, 517–529.

Holt, D., and Smith, T.M.F. (1979). Post Stratification. *Journal of the Royal Statistical Society*. Series A (General), 142(1), 33-46.

Kalton, G., and Flores Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, 19(2), 81-97.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, 29, 119-127.

Kish, L. (1992). Weighting for unequal pi. *Journal of Official Statistic,* 8, 183-200.

Kish, L. (1965). *Survey sampling*. New York: John Wiley and Sons.

Lepkowski, J., and Bowles, J. (1996). Sampling error software for personal computers. *The Survey Statistician*, 35, 10-16.

Liu, B., Ferraro, D., Wilson, E., and Brick, J.M. (2004). Trimming extreme weights in household surveys. *Proceedings of the Survey Research Methods Section of the American Statistical Association* [CD-ROM]. Alexandria, VA: American Statistical Association.

Office of Management and Budget (1997). *Revisions to the standards for the classification of federal data on race and ethnicity*, 62(210).

Oh, H.L., and Scheuren, F.J. (1983). *Weighting adjustments for unit nonresponse*. Incomplete data in sample surveys: *Vol. 2.* W.G. Madow, I. Olkin, and D.B. Rubin (Eds.). Academic Press, 143-184.

Research Triangle Institute (2009). *SUDAAN user's manual*, *Release 10.*. Research Triangle Park, NC: Research Triangle Institute.

Rizzo, Lou, Brick J.M., and Park, I. (2004). A minimally intrusive method for sampling persons in random-digit-dial surveys. *Public Opinion Quarterly,* 68, 267-274.

Rust, K.F., and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5, 282-310.

Sande, I. G. (1983). *Hot-deck imputation procedures*, *incomplete data in sample surveys*: *Vol. 3*. New York: Academic Press.

SAS Institute (2004). *SAS/STAT→ 9.1 User's Guide ,* Cary, NC: SAS Institute, Inc.

Shao, J. (1996). Resampling methods in sample surveys (with discussion). *Statistics*, 27, 203-254.

Skinner, C.J., Holt D., and Smith, T. M. F. (1989). *Analysis of complex surveys*. New York: John Wiley and Sons.

StataCorp (2009). *Stata Survey Data Reference Manual: Release 11*. College Station, TX: StataCorp LP.

State of California, Department of Finance (2004). *Population projections by race/ethnicity for California and it's counties, 2000-2050.* Sacramento, CA.

State of California, Department of Finance (2006a). *California county population estimates and components of change by year, July 1, 2000–2005*. Sacramento, CA.

State of California, Department of Finance (2006b). *Estimated race/ethnic population with age and sex detail*, 2000–2004. Sacramento, CA.

U.S. Census Bureau (2001). *2000 Summary File 1* (Prepared by the U.S. Census Bureau). California.

U.S. Census Bureau (2002). *Census 2000 Modified Race Data*. (Prepared by the U.S. Census Bureau).

U.S. Census Bureau (2003). *2002 American Community Survey PUMS file* (Prepared by the U.S. Census Bureau). California.

Verma, V., Scott, C., and O'Muircheartaigh, C. (1980). Sample designs and sampling errors for the world fertility survey. *Journal of the Royal Statistical Society*, A 143, 431-473.

Westat (2000). *WesVar ™ 4.0 user's guide.* Westat, Rockville, MD.

Wolter, K. (1985). *Introduction to variance estimation*. Springer-Verlag.

Appendix A

Table A-1. CHIS 2009 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean surname list ,Korean and any other race but Vietnamese surname list, Vietnamese surname list ,Vietnamese and any other race but Korean surname list)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Korean and any other race but Vietnamese surname list			Vietnamese surname list			Vietnamese and any other race but Korean surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
1.12	Los Angeles, San Fernando SPA – High Density	118,200	2,808	42.106	943	301	2.910	1,114	350	2.963	142	107	1.268	82	65	1.242
1.13	Los Angeles, San Gabriel SPA – High Density	368,000	16,143	22.500	1,627	510	2.890	10,163	3,198	2.838	4,199	3,112	1.301	2,188	1,727	1.235
1.14	Los Angeles, Metro SPA – High Density	406,800	16,781	24.176	2,655	830	2.870	3,039	962	2.843	264	201	1.282	175	143	1.215
1.17	Los Angeles, South SPA – High Density	115,400	5,885	19.499	804	261	2.851	1,436	437	2.838	277	196	1.351	157	118	1.276
1.18	Los Angeles, South Bay SPA – High Density	214,300	8,780	24.405	700	214	2.857	1,114	352	2.806	596	434	1.304	180	138	1.250
1.21	Los Angeles, Antelope Valley SPA – Low Density	224,100	7,291	30.724	172	57	2.820	171	53	2.850	222	164	1.306	51	41	1.244
1.22	Los Angeles, San Fernando SPA – Low Density	1,716,300	36,917	46.513	2,084	656	2.969	2,811	873	2.978	1,856	1,369	1.311	530	419	1.235
1.23	Los Angeles, San Gabriel SPA – Low Density	951,900	16,453	57.604	1,380	448	2.961	7,411	2,380	2.983	4,168	3,102	1.324	1,723	1,383	1.233
1.24	Los Angeles, Metro SPA – Low Density	643,100	13,120	49.027	1,168	376	2.920	2,239	714	2.962	931	682	1.321	475	376	1.224
1.25	Los Angeles, West SPA – Low Density	1,041,700	26,418	39.402	1,028	322	2.988	2,136	688	2.886	557	412	1.336	372	292	1.253
1.26	Los Angeles, South SPA – Low Density	703,600	18,644	37.709	678	216	2.922	819	264	2.957	275	200	1.341	68	51	1.333
1.27	Los Angeles, East SPA – Low Density	776,200	25,359	30.603	763	237	3.016	1,447	453	2.843	574	432	1.278	231	181	1.242
1.28	Los Angeles, South Bay SPA – Low Density	1,063,800	19,182	55.470	1,633	530	2.906	2,480	785	3.021	1,701	1,249	1.323	566	450	1.220
2.1	San Diego – High Density	366,100	27,175	13.397	265	90	2.500	585	178	2.813	2,702	1,988	1.301	394	311	1.235
2.2	San Diego – Low Density	2,111,000	84,290	25.021	1,334	410	2.925	2,524	808	2.814	2,417	1,795	1.308	626	493	1.245
3.1	Orange – High Density	1,009,600	34,484	29.133	3,314	1,036	2.920	5,027	1,601	2.871	15,631	11,505	1.309	2,563	2,023	1.233
3.2	Orange – Low Density	1,680,800	19,091	87.986	1,780	580	2.972	3,382	1,086	3.017	4,751	3,560	1.317	999	789	1.249
4.1	Santa Clara - High	566,300	10,600	53.464	758	241	3.081	3,144	1,001	2.947	7,367	5,421	1.323	1,568	1,242	1.232
4.2	Santa Clara - Low	1,042,300	16,955	61.494	1,909	603	3.016	8,118	2,544	3.019	7,585	5,649	1.316	2,109	1,665	1.241
5	San Bernardino	1,312,600	24,710	53.085	1,247	398	2.955	2,304	734	2.977	1,948	1,447	1.323	582	457	1.249
6	Riverside	1,350,800	26,461	51.046	1,133	360	2.935	1,603	504	2.985	2,136	1,584	1.318	472	375	1.236
7	Alameda	1,547,500	24,043	64.224	2,268	721	3.004	10,790	3,448	3.006	5,207	3,880	1.321	2,764	2,198	1.237
8	Sacramento	1,174,400	19,850	59.135	974	303	3.006	2,800	893	2.972	3,555	2,646	1.319	1,056	836	1.244
9	Contra Costa	953,300	16,587	57.382	820	266	3.026	2,673	843	3.031	1,266	943	1.317	643	514	1.237

Table A-1. CHIS 2009 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean surname list ,Korean and any other race but Vietnamese surname list, Vietnamese surname list ,Vietnamese and any other race but Korean surname list) (Continued)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Korean and any other race but Vietnamese surname list			Vietnamese surname list			Vietnamese and any other race but Korean surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
10	Fresno	669,400	11,400	58.697	493	160	2.952	1,007	325	2.988	807	598	1.314	180	143	1.233
11	San Francisco	1,124,600	27,269	41.004	1,773	565	2.935	12,694	4,082	2.936	3,108	2,325	1.315	3,189	2,550	1.236
12	Ventura	634,400	15,972	39.736	422	137	2.890	845	272	2.914	706	509	1.307	185	147	1.233
13	San Mateo	826,100	13,266	62.235	1,029	331	2.965	4,429	1,415	3.003	783	583	1.309	1,029	812	1.240
14	Kern	534,700	9,100	58.767	278	85	3.089	284	87	3.087	353	265	1.303	62	49	1.216
15	San Joaquin	422,900	7,999	52.819	353	106	3.070	674	216	2.996	1,129	841	1.319	409	327	1.221
16	Sonoma	460,200	7,738	59.472	318	101	3.029	344	109	2.940	432	321	1.321	105	84	1.221
17	Stanislaus	325,200	7,170	45.310	187	61	2.968	224	71	3.068	309	232	1.304	58	45	1.261
18	Santa Barbara	381,700	11,476	33.307	216	64	2.769	320	98	2.883	269	191	1.325	63	49	1.212
19	Solano	310,300	9,214	33.687	162	55	2.746	361	112	2.935	391	286	1.295	107	85	1.230
20	Tulare	286,500	8,749	32.738	76	26	2.621	142	47	2.840	135	98	1.337	20	16	1.250
21	Santa Cruz	265,500	7,502	35.393	133	42	2.771	228	70	2.961	147	106	1.349	38	31	1.226
22	Marin	332,800	37,763	8.809	228	66	2.682	400	114	2.581	333	222	1.316	87	58	1.225
23	San Luis Obispo	246,500	6,230	39.537	119	39	3.051	111	36	2.775	154	116	1.305	36	29	1.241
24	Placer	306,500	7,288	42.026	201	62	3.000	294	89	3.128	294	209	1.374	53	44	1.205
25	Merced	124,900	6,700	18.610	94	32	2.848	144	45	2.939	124	91	1.319	21	16	1.235
26	Butte	162,400	5,550	29.270	112	36	2.732	111	36	2.775	176	128	1.323	49	36	1.256
27	Shasta	143,800	5,610	25.633	93	31	2.818	45	14	2.813	79	55	1.339	12	8	1.333
28	Yolo	145,900	7,726	18.850	164	54	2.485	551	170	2.840	256	189	1.313	107	84	1.259
29	El Dorado	147,700	7,140	20.684	93	31	2.657	124	39	2.531	100	78	1.266	27	20	1.286
30	Imperial	89,600	9,521	9.408	37	11	3.083	77	22	2.406	31	22	1.292	7	5	1.400
31	Napa	122,100	7,960	15.338	61	19	2.542	55	17	2.619	58	41	1.318	17	13	1.308
32	Kings	72,600	7,307	9.937	54	15	2.700	41	12	2.278	54	39	1.227	4	3	1.333
33	Madera	88,400	8,019	11.021	36	9	2.769	50	14	2.632	44	33	1.294	5	4	1.250
34	Monterey	348,800	8,258	42.182	265	84	2.978	369	121	2.906	269	201	1.332	64	51	1.255
35	Humboldt	119,100	8,785	13.552	58	19	2.762	49	13	2.882	64	47	1.306	10	7	1.111
36	Nevada	95,700	6,671	14.333	74	23	2.846	61	20	2.773	65	49	1.275	11	8	1.100
37	Mendocino	86,700	7,467	11.606	36	10	3.273	51	16	2.550	67	45	1.340	13	9	1.300
38	Sutter	53,100	6,450	8.229	37	14	1.947	39	15	1.950	51	37	1.214	31	24	1.240

Table A-1. CHIS 2009 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean surname list ,Korean and any other race but Vietnamese surname list, Vietnamese surname list ,Vietnamese and any other race but Korean surname list) (Continued)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Korean and any other race but Vietnamese surname list			Vietnamese surname list			Vietnamese and any other race but Korean surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
39	Yuba	55,900	7,257	7.696	49	12	2.722	64	20	2.560	48	34	1.231	17	13	1.214
40	Lake	65,300	7,866	8.305	48	13	2.000	22	7	2.750	24	18	1.200	4	2	1.333
41	San Benito	42,400	9,551	4.437	14	3	2.800	15	5	1.667	18	15	1.200	4	3	1.333
42	Colusa, Glenn, Tehama	87,200	3,950	22.063	32	14	2.286	27	9	3.000	47	33	1.343	10	8	1.111
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	177,300	5,809	30.498	78	21	3.545	38	12	2.923	107	85	1.244	5	5	1.000
44	Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne	234,800	5,020	46.768	119	38	2.833	93	26	3.100	172	128	1.344	17	12	1.417
	Total	31,049,100	848,780		38,979	12,355		103,713	32,925		81,531	60,348		26,630	21,087	

[1] Total number of possible phone numbers in eligible working 100 banks
[2] Realized number of sampled telephone numbers in strata.

Table A-2. CHIS 2009 cell-phone sample frame size, sample sizes, and base weights by sampling stratum or area code

Sampling stratum (Area code)	Number of cell phone numbers	Number of sampled cell phone numbers	Weight
Total	45,428,000	58,900	
209	1,825,000	3,500	521.43
213	1,197,000	1,600	748.13
310	2,349,000	700	3,355.71
323	2,624,000	1,800	1,457.78
408	1,956,000	1,200	1,630.00
415	1,549,000	1,200	1,290.83
510	2,035,000	600	3,391.67
530	1,446,000	13,200	109.55
559	1,624,000	5,700	284.91
562	1,828,000	4,400	415.45
619	2,311,000	600	3,851.67
626	1,748,000	600	2,913.33
650	1,137,000	600	1,895.00
661	1,380,000	500	2,760.00
707	1,542,000	3,900	395.38
714	2,655,000	600	4,425.00
760	2,263,000	900	2,514.44
805	1,923,000	3,600	534.17
818	2,343,000	1,100	2,130.00
831	802,000	3,800	211.05
858	754,000	900	837.78
909	2,093,000	1,200	1,744.17
916	1,965,000	3,300	595.45
925	1,213,000	1,200	1,010.83
949	1,131,000	1,400	807.86
951	1,735,000	800	2,168.75

Appendix B

Table B-1. Household weighting procedures by sample type

	All Samples	Landline	List	Cell
1. Base weight				
1.1 Sample size	1,034,395	848,780	126,715	58,900
1.2 Sum of weights	76,477,100	30,810,610	238,490	45,428,000
1.3 Coefficient of variation		51.84	42.00	116.48
2. Adjusting for unreleased landline cell phone cases				
2.1 Sum of weights before adjustment				
a. Not a landline cell case	76,264,271	30,599,601	236,670	45,428,000
b. Released landline cell cases	154,778	154,778	0	0
c. Unreleased landline cell cases	58,051	56,231	1,820	0
2.2 Sum of weights after adjustment				
a. Not a landline cell case	76,264,271	30,599,601	236,670	45,428,000
b. Released landline cell cases	212,829	0	0	212,829
c. Unreleased landline cell cases	0	0	0	0
2.3 Sample size	1,034,395	843,035	125,697	65,663
2.4 Coefficient of variation		51.87	41.99	126.62
3. CATI extraction and adjusting for new work subsampling				
3.1 Sum of weights before adjustment				
a. Not subsampled	76,035,038	30,224,807	169,403	45,640,829
b. Subsampled	442,062	374,794	67,267	0
3.2 Sum of weights after adjustment				
a. Not subsampled	76,477,100	30,596,556	236,670	45,643,874
b. Subsampled	0	0	0	0
3.3 Sample size	984,107	831,663	89,670	62,774
3.4 Coefficient of variation		51.66	41.14	122.00

Table B-1. Household weighting procedures by sample type (continued)

				All Samples	Landline	List	Cell
4.	First refusal conversion subsampling adjustment						
	4.1	Sum of weights before adjustment					
		a.	Household never refused	58,017,101	26,329,877	195,689	31,491,534
		b.	Household refused - selected for refusal conversion	18,459,999	4,266,679	40,981	14,152,340
		c.	Household refused - not selected for refusal conversion	0	0	0	0
	4.2	Sum of weights after adjustment					
		a.	Household never refused	58,017,101	26,329,877	195,689	31,491,534
		b.	Household refused - selected for refusal conversion	18,459,999	4,266,679	40,981	14,152,340
		c.	Household refused - not selected for refusal conversion	0	0	0	0
	4.3	Sample size		984,107	831,663	89,670	62,774
	4.4	Coefficient of variation			51.66	41.14	122.00
5.	Second refusal conversion subsampling						
	5.1	Sum of weights before adjustment					
		a.	Household never refused more than once	60,761,513	27,385,936	207,703	33,167,874
		b.	Household refused -selected for second refusal conversion	15,608,480	3,103,513	28,967	12,476,001
		c.	Household refused -not selected for second refusal conversion	107,107	107,107	0	0
	5.2	Sum of weights after adjustment					
		a.	Household never refused more than once	60,761,513	27,385,936	207,703	33,167,874
		b.	Household refused -selected for second refusal conversion	15,715,587	3,210,620	28,967	12,476,001
		c.	Household refused -not selected for second refusal conversion	0	0	0	0
	5.3	Sample size		980,292	827,848	89,670	62,774
	5.4	Coefficient of variation			51.53	41.14	122.00

Table B-1. Household weighting procedures by sample type (continued)

	All Samples	Landline	List	Cell
6. Adjusting for unknown residential status				
6.1 Sum of weights before adjustment				
a. Residential - respondents	8,758,924	2,916,054	22,601	5,820,268
b. Residential - nonrespondents	18,528,792	3,653,949	38,479	14,836,365
c. Unknown residential status (NA, NM)	15,155,366	4,075,307	12,027	11,068,032
d. Nonresidential	34,034,018	19,951,246	163,563	13,919,210
6.2 Sum of weights - allocating unknown residential				
a. Residential - respondents	8,758,924	2,916,054	22,601	5,820,268
b. Residential - nonrespondents	18,528,792	3,653,949	38,479	14,836,365
c. (NA, NM)	8,157,534	1,508,574	3,768	6,645,192
6.3 Sum of weights after adjustment				
a. Residential - respondents	8,758,924	2,916,054	22,601	5,820,268
b. Residential - nonrespondents	26,686,326	5,162,523	42,246	21,481,557
c. Estimated residential among unknown	0	0	0	0
6.4 Sample size	228,838	178,843	22,638	27,357
6.5 Coefficient of variation		57.55	42.21	122.95
7. Supplemental list-sample eligibility adjustment				
7.1 Sum of weights before adjustment				
a. Cell sample or not Korean or Vietnamese	35,028,119	7,726,294	0	27,301,825
b. Completed Korean or Vietnamese	362,710	352,284	10,426	0
c. Nonresponse, but known that it is not Korean or Vietnamese	12,296	0	12,296	0
d. Nonresponse, unknown Korean or Vietnamese status	42,125	0	42,125	0
7.2 Sum of weights after adjustment				
a. Cell sample or not Korean or Vietnamese	35,061,052	7,759,227	0	27,301,825
b. Completed Korean or Vietnamese	384,198	353,576	30,622	0
c. Nonresponse, but known that it is not Korean or Vietnamese	0	0	0	0
d. Nonresponse, unknown Korean or Vietnamese status	0	0	0	0
7.3 Sample size	210,324	178,843	4,124	27,357
7.4 Coefficient of variation		57.63	27.19	122.95

Table B-1. Household weighting procedures by sample type (continued)

			All Samples	Landline	List	Cell
8.	Unknown presence of children in household					
	8.1	Sum of weights before adjustment				
	a.	Ineligible respondent	2,138,605	7,735	0	2,130,870
	b.	Eligible respondent - child status known	6,639,820	2,920,314	30,109	3,689,398
	c.	Eligible nonrespondent - child status known	56,054	36,557	309	19,188
	d.	Unknown nonrespondent - child status unknown	26,610,771	5,148,197	204	21,462,369
	8.2	Sum of weights after adjustment				
	a.	Ineligible respondent	2,138,605	7,735	0	2,130,870
	b.	Eligible respondent - child status known	6,695,875	2,956,871	30,418	3,708,586
	c.	Eligible nonrespondent - child status known	0	0	0	0
	d.	Unknown nonrespondent - child status unknown	26,610,771	5,148,197	204	21,462,369
	8.3	Sample Size	209,576	178,142	4,089	27,345
	8.4	Coefficient of variation		57.42	27.28	122.93
9.	Screener nonresponse adjustment					
	9.1	Sum of weights before adjustment				
	a.	Respondents	8,834,479	2,964,606	30,418	5,839,456
	b.	Nonrespondents	26,610,771	5,148,197	204	21,462,369
	9.2	Sum of weights after adjustment				
	a.	Respondents	35,445,250	8,112,803	30,622	27,301,825
	b.	Nonrespondents	0	0	0	0
	9.3	Sample size	93,855	81,611	4,061	8,183
	9.4	Coefficient of variation		56.09	27.28	129.47
10.	Multiple telephone adjustment					
	10.1	Sum of weights before adjustment	25,462,659	8,089,676	30,622	17,342,361
	10.2	Sum of weights after adjustment	25,276,257	7,903,290	30,606	17,342,361
	10.3	Sample size	778,824	769,560	4,061	5,203
	10.4	Coefficient of variation		57.27	27.29	129.91
	10.5	Overall adjustment factor	99.3%	97.7%	99.9%	100.0%

Table B-1. Household weighting procedures by sample type (continued)

		All Samples	Landline	List	Cell
11.	Duplicate respondent adjustment				
11.1	Sum of weights before adjustment				
	a. Not a duplicate number	25,253,103	7,902,897	30,601	17,319,605
	b. Duplicate number	23,153	393	5	22,755
11.2	Sum of weights after adjustment				
	a. Not a duplicate number	25,276,257	7,903,290	30,606	17,342,361
	b. Duplicate number	0	0	0	0
11.3	Sample size	90,631	81,375	4,060	5,196
11.4	Coefficient of variation		57.25	27.82	129.87
12.	Section G nonresponse adjustment*				
12.1	Sum of weights before adjustment				
	a. Household with child 1st procedure	562,607	559,424	3,182	0
	b. Household w/o child 1st procedure - section G completed	14,288,918	4,037,496	13,974	10,237,449
	c. Household w/o child 1st procedure - section G not completed	10,424,731	3,306,370	13,450	7,104,912
12.2	Sum of weights after adjustment				
	a. Household with child 1st procedure	562,607	559,424	3,182	0
	b. Household w/o child 1st procedure - section G completed	24,713,650	7,343,866	27,424	17,342,361
	c. Household w/o child 1st procedure - section G not completed	0	0	0	0
12.3	Sample size	52,748	47,370	2,260	3,118
12.4	Coefficient of variation		61.63	35.32	129.81

* Adjustment applicable to the child and adolescent samples. See Section 5.1.

Table B-2. Extended interview weighting procedures for adult interviews by sample type

			All Samples	RDD	List	Cell
1.	Adult Base Weight					
	1.1	Number of Sampled Adults	90,631	81,375	4,060	5,196
	1.2	Sum of Weights	35,996,242	16,630,278	73,267	19,292,696
	1.3	Coefficient of Variation		77.52	47.46	143.02
2.	Nonresponse Adjustment					
	2.1	Sum of Weights Before Adjustment				
		a. Eligible Respondents	18,779,706	8,016,388	33,600	10,729,718
		b. Ineligible	503,922	289,397	2,146	212,379
		c. Nonrespondents	16,712,614	8,324,493	37,522	8,350,599
	2.2	Sum of Weights After Adjustment				
		a. Eligible Respondents	35,019,500	16,039,972	68,857	18,910,671
		b. Ineligible	976,742	590,306	4,411	382,025
		c. Nonrespondents	0	0	0	0
	2.3	Number of Completed Interviews	47,614	42,682	1,885	3,047
	2.4	Coefficient of Variation (CV)		78.31	45.91	133.93
	2.5	Mean Adjustment Factor	1.92	2.00	2.05	1.80

Table B-3. Extended interview weighting procedures for child interviews by sample type

	All Samples	RDD	List	Cell
1. Child Base Weight				
1.1 Number of Sampled Children	12,468	11,057	791	620
1.2 Sum of Weights	12,054,881	3,473,871	16,108	8,564,902
1.3 Coefficient of Variation		112.66	73.23	195.37
2. Nonresponse Adjustment				
2.1 Sum of Weights Before Adjustment				
a. Eligible Respondents	9,012,483	2,508,905	11,349	6,492,229
b. Ineligible	40,434	18,854	111	21,470
c. Nonrespondents	3,001,964	946,111	4,649	2,051,204
2.2 Sum of Weights After Adjustment				
a. Eligible Respondents	12,002,219	3,448,478	15,943	8,537,799
b. Ineligible	52,662	25,393	166	27,103
c. Nonrespondents	0	0	0	0
2.3 Number of Completed Interviews	8,945	7,918	545	482
2.4 Coefficient of Variation (CV)		108.61	45.91	195.64
2.5 Mean Adjustment Factor	1.34	1.37	1.41	1.32

Table B-4. Extended interview weighting adjustments for adolescent interviews by sample type

	All Samples	RDD	List	Cell
1. Teen Base Weight				
1.1 Number of Sampled Children	8,030	7,121	500	409
1.2 Sum of Weights	6,644,532	2,106,849	9,928	4,527,755
1.3 Coefficient of Variation		96.53	70.53	156.88
2. Nonresponse Adjustment				
2.1 Sum of Weights Before Adjustment				
a. Eligible Respondents	2,974,658	892,940	3,443	2,078,274
b. Ineligible	72,341	26,900	186	45,255
c. Nonrespondents	3,597,533	1,187,009	6,299	2,404,225
2.2 Sum of Weights After Adjustment				
a. Eligible Respondents	6,502,472	2,044,215	9,474	4,448,783
b. Ineligible	142,059	62,634	454	78,972
c. Nonrespondents	0	0	0	0
2.3 Number of Completed Interviews	3,379	3,002	178	199
2.4 Coefficient of Variation (CV)		96.73	66.06	159.49
2.5 Mean Adjustment Factor	2.23	2.29	2.75	2.18

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum

			All Strata	Los Angeles	San Diego	Orange	Santa Clara
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	47,614	9,148	5,122	2,697	1,731
	1.2	Sum of weights before poststratification	35,019,500	8,770,453	3,017,425	2,531,544	1,608,714
	1.3	Sum of weights after poststratification	45,333,122	11,235,076	3,983,853	3,253,852	2,142,047
2.	Composite weight						
	2.1 Sum of weights after composite factor		27,546,653	6,945,561	2,263,809	2,137,303	1,338,236
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	70	10	21	0	0
	3.2	Sum of Weights Before Trimming Adjustment	27,546,653	6,945,561	2,263,809	2,137,303	1,338,236
	3.3	Sum of Weights After Trimming Adjustment	26,947,821	6,831,676	2,016,557	2,137,303	1,338,236
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	47,614	9,148	5,122	2,697	1,731
	4.2	Sum of Weights After Adjustment	27,546,653	7,446,871	2,275,695	2,301,703	1,365,232
	4.3	Coefficient of Variation (CV)	235.64	215.56	194.24	264.59	194.50
	4.4	Mean Adjustment Factor	1.02	1.09	1.13	1.08	1.02
	4.5	Mean Weight	578.54	814.04	444.30	853.43	788.70

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum (continued)

			San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	1,515	1,719	1,221	1,403	1,008
	1.2	Sum of weights before poststratification	1,609,266	2,348,975	1,611,778	1,404,963	1,080,480
	1.3	Sum of weights after poststratification	2,099,034	2,964,291	2,108,112	1,811,893	1,444,431
2.	Composite weight						
	2.1 Sum of weights after composite factor		1,281,664	1,723,752	1,194,406	1,076,873	854,031
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	3	0	0	0
	3.2	Sum of Weights Before Trimming Adjustment	1,281,664	1,723,752	1,194,406	1,076,873	854,031
	3.3	Sum of Weights After Trimming Adjustment	1,281,664	1,674,206	1,194,406	1,076,873	854,031
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	1,515	1,719	1,221	1,403	1,008
	4.2	Sum of Weights After Adjustment	1,421,349	1,469,418	1,163,885	1,030,148	797,068
	4.3	Coefficient of Variation (CV)	216.18	246.28	211.55	135.98	180.15
	4.4	Mean Adjustment Factor	1.11	0.88	0.97	0.96	0.93
	4.5	Mean Weight	938.18	854.81	953.22	734.25	790.74

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum (continued)

			Fresno	San Francisco	Ventura	San Mateo	Kern
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	829	809	1,007	586	601
	1.2	Sum of weights before poststratification	762,576	961,372	757,381	685,208	662,025
	1.3	Sum of weights after poststratification	967,399	1,213,105	1,011,769	909,298	815,859
2.	Composite weight						
	2.1 Sum of weights after composite factor		611,841	720,985	602,568	583,017	549,941
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	3	2	0	0
	3.2	Sum of Weights Before Trimming Adjustment	611,841	720,985	602,568	583,017	549,941
	3.3	Sum of Weights After Trimming Adjustment	611,841	700,564	595,993	583,017	549,941
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	829	809	1,007	586	601
	4.2	Sum of Weights After Adjustment	650,509	702,799	609,236	569,380	552,895
	4.3	Coefficient of Variation (CV)	120.73	215.46	140.93	189.97	209.73
	4.4	Mean Adjustment Factor	1.06	1.00	1.02	0.98	1.01
	4.5	Mean Weight	784.69	868.73	605.00	971.64	919.96

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum (continued)

			San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	588	583	515	659	546
	1.2	Sum of weights before poststratification	662,042	478,627	455,687	404,401	465,593
	1.3	Sum of weights after poststratification	861,778	627,534	609,610	506,299	585,393
2.	Composite weight						
	2.1	Sum of weights after composite factor	482,350	402,033	325,388	322,028	372,037
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	2	0	0	3
	3.2	Sum of Weights Before Trimming Adjustment	482,350	402,033	325,388	322,028	372,037
	3.3	Sum of Weights After Trimming Adjustment	482,350	386,986	325,388	322,028	354,091
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	588	583	515	659	546
	4.2	Sum of Weights After Adjustment	448,114	359,619	352,209	306,262	301,356
	4.3	Coefficient of Variation (CV)	132.73	138.00	123.89	156.77	145.54
	4.4	Mean Adjustment Factor	0.93	0.93	1.08	0.95	0.85
	4.5	Mean Weight	762.10	616.84	683.90	464.74	551.93

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum (continued)

			Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	569	559	2,063	517	570
	1.2	Sum of weights before poststratification	390,592	285,342	227,578	274,099	350,860
	1.3	Sum of weights after poststratification	474,475	379,257	316,969	356,317	479,086
2.	Composite weight						
	2.1	Sum of weights after composite factor	313,771	228,500	201,698	230,774	283,363
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	1	1	2	1	0
	3.2	Sum of Weights Before Trimming Adjustment	313,771	228,500	201,698	230,774	283,363
	3.3	Sum of Weights After Trimming Adjustment	296,404	222,608	196,708	223,285	283,363
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	569	559	2,063	517	570
	4.2	Sum of Weights After Adjustment	297,992	201,835	191,257	197,562	252,553
	4.3	Coefficient of Variation (CV)	109.62	124.32	224.64	161.68	112.43
	4.4	Mean Adjustment Factor	1.01	0.91	0.97	0.89	0.89
	4.5	Mean Weight	523.71	361.07	92.71	382.13	443.08

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum (continued)

			Merced	Butte	Shasta	Yolo	El Dorado
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	537	626	602	617	561
	1.2	Sum of weights before poststratification	204,848	225,842	188,141	255,797	175,279
	1.3	Sum of weights after poststratification	271,318	281,120	249,871	324,677	238,510
2.	Composite weight						
	2.1	Sum of weights after composite factor	155,860	186,185	147,274	182,065	153,552
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	0	0	0	1
	3.2	Sum of Weights Before Trimming Adjustment	155,860	186,185	147,274	182,065	153,552
	3.3	Sum of Weights After Trimming Adjustment	155,860	186,185	147,274	182,065	150,582
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	537	626	602	617	561
	4.2	Sum of Weights After Adjustment	175,897	167,285	138,318	144,614	140,981
	4.3	Coefficient of Variation (CV)	173.46	131.77	122.57	196.09	103.83
	4.4	Mean Adjustment Factor	1.13	0.90	0.94	0.79	0.94
	4.5	Mean Weight	327.55	267.23	229.76	234.38	251.30

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum (continued)

			Imperial	Napa	Kings	Madera	Monterey
1.		Poststratification to telephone service					
	1.1	Number of Completed Interviews	537	514	497	556	598
	1.2	Sum of weights before poststratification	178,880	137,285	117,511	119,265	379,041
	1.3	Sum of weights after poststratification	239,840	176,123	156,185	159,253	489,078
2.		Composite weight					
	2.1	Sum of weights after composite factor	109,068	120,641	86,671	92,296	305,713
3.		Trimming Adjustment*					
	3.1	Number of Trimmed Records	4	0	0	0	0
	3.2	Sum of Weights Before Trimming Adjustment	109,068	120,641	86,671	92,296	305,713
	3.3	Sum of Weights After Trimming Adjustment	82,777	120,641	86,671	92,296	305,713
4.		Raking Adjustment*					
	4.1	Number of Completed Interviews	537	514	497	556	598
	4.2	Sum of Weights After Adjustment	125,706	97,229	94,585	101,557	296,059
	4.3	Coefficient of Variation (CV)	143.37	211.73	160.99	136.48	113.00
	4.4	Mean Adjustment Factor	1.52	0.81	1.09	1.10	0.97
	4.5	Mean Weight	234.09	189.16	190.31	182.66	495.08

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum (continued)

			Humboldt	Nevada	Mendocino	Sutter	Yuba
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	876	581	593	529	455
	1.2	Sum of weights before poststratification	176,272	161,738	89,409	87,527	59,423
	1.3	Sum of weights after poststratification	217,385	232,069	121,878	116,626	74,137
2.	Composite weight						
	2.1	Sum of weights after composite factor	132,319	104,425	69,951	62,162	51,114
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	1	2	3	0	0
	3.2	Sum of Weights Before Trimming Adjustment	132,319	104,425	69,951	62,162	51,114
	3.3	Sum of Weights After Trimming Adjustment	119,550	100,657	65,634	62,162	51,114
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	876	581	593	529	455
	4.2	Sum of Weights After Adjustment	101,238	80,081	68,066	65,789	49,525
	4.3	Coefficient of Variation (CV)	263.46	104.99	120.40	117.66	170.16
	4.4	Mean Adjustment Factor	0.85	0.80	1.04	1.06	0.97
	4.5	Mean Weight	115.57	137.83	114.78	124.36	108.85

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-5. Poststratification, composite weighting, trimming, and raking adjustments for adult interviews by self-reported stratum (continued)

			Lake	San Benito	Tehama, Etc.	Del Norte, Etc.	Tuolumnelaveras, Etc.
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	525	526	420	498	401
	1.2	Sum of weights before poststratification	50,869	68,255	121,186	210,592	205,355
	1.3	Sum of weights after poststratification	61,107	88,127	155,508	254,566	269,006
2.	Composite weight						
	2.1	Sum of weights after composite factor	52,772	47,364	94,735	159,688	186,869
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	4	1	2	2	1
	3.2	Sum of Weights Before Trimming Adjustment	52,772	47,364	94,735	159,688	186,869
	3.3	Sum of Weights After Trimming Adjustment	47,171	46,849	92,543	138,288	174,270
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	525	526	420	498	401
	4.2	Sum of Weights After Adjustment	50,579	41,017	83,093	112,482	147,608
	4.3	Coefficient of Variation (CV)	100.98	217.34	122.25	132.51	99.84
	4.4	Mean Adjustment Factor	1.07	0.88	0.90	0.81	0.85
	4.5	Mean Weight	96.34	77.98	197.84	225.87	368.10

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum

			All Strata	Los Angeles	San Diego	Orange	Santa Clara
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	8,945	1,755	949	538	434
	1.2	Sum of weights before poststratification	12,002,219	3,175,409	656,611	1,087,936	415,538
	1.3	Sum of weights after poststratification	10,238,601	2,634,823	672,580	828,116	420,910
2.	Composite weight						
	2.1	Sum of weights after composite factor	6,401,339	1,688,257	409,650	602,755	276,850
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	71	22	6	3	1
	3.2	Sum of Weights Before Trimming Adjustment	6,401,339	1,688,257	409,650	602,755	276,850
	3.3	Sum of Weights After Trimming Adjustment	5,549,723	1,445,812	386,356	460,068	261,029
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	8,945	1,755	949	538	434
	4.2	Sum of Weights After Adjustment	6,398,447	1,728,356	530,528	514,212	308,054
	4.3	Coefficient of Variation (CV)	170.57	167.41	111.55	214.89	142.30
	4.4	Mean Adjustment Factor	1.15	1.20	1.37	1.12	1.18
	4.5	Mean Weight	715.31	984.82	559.04	955.78	709.80

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum (continued)

			San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	321	312	259	247	173
	1.2	Sum of weights before poststratification	504,126	973,418	568,483	396,804	276,433
	1.3	Sum of weights after poststratification	443,300	712,117	530,138	374,421	256,863
2.	Composite weight						
	2.1	Sum of weights after composite factor	331,790	482,806	227,813	209,414	178,471
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	5	5	3	0	3
	3.2	Sum of Weights Before Trimming Adjustment	331,790	482,806	227,813	209,414	178,471
	3.3	Sum of Weights After Trimming Adjustment	290,618	281,802	221,668	209,414	155,882
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	321	312	259	247	173
	4.2	Sum of Weights After Adjustment	374,875	377,005	244,339	243,667	161,806
	4.3	Coefficient of Variation (CV)	124.72	131.11	115.77	118.88	119.96
	4.4	Mean Adjustment Factor	1.29	1.34	1.10	1.16	1.04
	4.5	Mean Weight	1167.83	1208.35	943.39	986.51	935.29

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum (continued)

			Fresno	San Francisco	Ventura	San Mateo	Kern
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	192	110	189	106	142
	1.2	Sum of weights before poststratification	379,935	327,812	240,186	259,917	343,140
	1.3	Sum of weights after poststratification	307,004	278,359	225,313	201,847	243,007
2.	Composite weight						
	2.1	Sum of weights after composite factor	187,456	121,561	118,934	162,280	197,205
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	1	0	3	3
	3.2	Sum of Weights Before Trimming Adjustment	187,456	121,561	118,934	162,280	197,205
	3.3	Sum of Weights After Trimming Adjustment	187,456	116,349	118,934	131,257	141,965
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	192	110	189	106	142
	4.2	Sum of Weights After Adjustment	179,514	86,985	138,458	117,444	165,483
	4.3	Coefficient of Variation (CV)	109.71	151.99	108.49	135.82	135.42
	4.4	Mean Adjustment Factor	0.96	0.75	1.16	0.90	1.17
	4.5	Mean Weight	934.97	790.78	732.58	1107.97	1165.37

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum (continued)

			San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	127	104	105	97	88
	1.2	Sum of weights before poststratification	614,090	100,762	134,743	129,223	122,249
	1.3	Sum of weights after poststratification	508,452	97,118	120,059	109,248	100,705
2.	Composite weight						
	2.1	Sum of weights after composite factor	153,998	70,338	82,620	64,242	72,411
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	1	0	2	0
	3.2	Sum of Weights Before Trimming Adjustment	153,998	70,338	82,620	64,242	72,411
	3.3	Sum of Weights After Trimming Adjustment	153,998	67,542	82,620	57,498	72,411
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	127	104	105	97	88
	4.2	Sum of Weights After Adjustment	151,894	76,960	106,356	69,086	68,311
	4.3	Coefficient of Variation (CV)	137.02	110.29	112.56	132.72	122.70
	4.4	Mean Adjustment Factor	0.99	1.14	1.29	1.20	0.94
	4.5	Mean Weight	1196.02	740.00	1012.91	712.23	776.26

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum (continued)

			Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	123	102	326	64	118
	1.2	Sum of weights before poststratification	131,711	56,834	39,655	39,315	92,900
	1.3	Sum of weights after poststratification	110,460	59,393	43,399	42,696	93,633
2.	Composite weight						
	2.1	Sum of weights after composite factor	80,973	36,785	39,718	27,896	57,708
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	0	1	0	0
	3.2	Sum of Weights Before Trimming Adjustment	80,973	36,785	39,718	27,896	57,708
	3.3	Sum of Weights After Trimming Adjustment	80,973	36,785	34,206	27,896	57,708
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	123	102	326	64	118
	4.2	Sum of Weights After Adjustment	88,122	42,137	38,838	32,597	52,985
	4.3	Coefficient of Variation (CV)	99.67	108.20	76.41	107.56	106.86
	4.4	Mean Adjustment Factor	1.09	1.15	1.14	1.17	0.92
	4.5	Mean Weight	716.44	413.11	119.14	509.33	449.02

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum (continued)

			Merced	Butte	Shasta	Yolo	El Dorado
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	117	88	90	125	79
	1.2	Sum of weights before poststratification	35,362	50,714	44,366	63,681	23,337
	1.3	Sum of weights after poststratification	39,910	49,015	40,478	61,279	29,318
2.	Composite weight						
	2.1	Sum of weights after composite factor	36,973	26,351	25,297	29,361	21,828
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	1	0	0	0	3
	3.2	Sum of Weights Before Trimming Adjustment	36,973	26,351	25,297	29,361	21,828
	3.3	Sum of Weights After Trimming Adjustment	35,034	26,351	25,297	29,361	21,128
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	117	88	90	125	79
	4.2	Sum of Weights After Adjustment	51,198	29,169	27,899	30,798	22,685
	4.3	Coefficient of Variation (CV)	88.55	84.24	92.52	102.91	78.00
	4.4	Mean Adjustment Factor	1.46	1.11	1.10	1.05	1.07
	4.5	Mean Weight	437.59	331.47	309.99	246.39	287.15

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum (continued)

			Imperial	Napa	Kings	Madera	Monterey
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	139	67	127	142	131
	1.2	Sum of weights before poststratification	119,796	10,759	44,839	49,202	156,848
	1.3	Sum of weights after poststratification	78,499	14,223	42,604	47,656	130,427
2.	Composite weight						
	2.1	Sum of weights after composite factor	58,128	13,033	21,852	25,423	87,582
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	1	0	0	0	0
	3.2	Sum of Weights Before Trimming Adjustment	58,128	13,033	21,852	25,423	87,582
	3.3	Sum of Weights After Trimming Adjustment	28,364	13,033	21,852	25,423	87,582
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	139	67	127	142	131
	4.2	Sum of Weights After Adjustment	30,038	21,068	31,516	27,617	82,028
	4.3	Coefficient of Variation (CV)	197.56	85.60	86.03	110.99	116.60
	4.4	Mean Adjustment Factor	1.06	1.62	1.44	1.09	0.94
	4.5	Mean Weight	216.10	314.44	248.16	194.48	626.17

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum (continued)

			Humboldt	Nevada	Mendocino	Sutter	Yuba
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	116	78	78	101	95
	1.2	Sum of weights before poststratification	42,688	30,761	14,375	42,467	16,933
	1.3	Sum of weights after poststratification	39,806	29,898	14,888	32,463	15,909
2.	Composite weight						
	2.1	Sum of weights after composite factor	13,736	15,156	7,955	20,323	11,481
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	0	0	2	0
	3.2	Sum of Weights Before Trimming Adjustment	13,736	15,156	7,955	20,323	11,481
	3.3	Sum of Weights After Trimming Adjustment	13,736	15,156	7,955	17,857	11,481
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	116	78	78	101	95
	4.2	Sum of Weights After Adjustment	17,984	10,513	11,881	19,369	12,114
	4.3	Coefficient of Variation (CV)	117.48	108.91	104.08	144.14	97.68
	4.4	Mean Adjustment Factor	1.31	0.69	1.49	1.09	1.06
	4.5	Mean Weight	155.04	134.78	152.31	191.77	127.52

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-6. Poststratification, composite weighting procedures, trimming and raking adjustments for child interview by self-reported stratum (continued)

			Lake	San Benito	Tehama, Etc.	Del Norte, Etc.	Tuolumnelaveras, Etc.
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	77	134	72	62	46
	1.2	Sum of weights before poststratification	18,971	17,379	41,799	33,022	77,692
	1.3	Sum of weights after poststratification	13,587	13,729	33,363	31,971	65,618
2.	Composite weight						
	2.1	Sum of weights after composite factor	13,005	11,289	22,943	19,562	38,128
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	1	3	0	0	1
	3.2	Sum of Weights Before Trimming Adjustment	13,005	11,289	22,943	19,562	38,128
	3.3	Sum of Weights After Trimming Adjustment	7,705	7,945	22,943	19,562	31,709
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	77	134	72	62	46
	4.2	Sum of Weights After Adjustment	7,569	10,797	16,116	21,462	18,616
	4.3	Coefficient of Variation (CV)	83.93	117.38	114.63	114.51	117.39
	4.4	Mean Adjustment Factor	0.98	1.36	0.70	1.10	0.59
	4.5	Mean Weight	98.30	80.57	223.83	346.16	404.70

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum

			All Strata	Los Angeles	San Diego	Orange	Santa Clara
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	3,379	659	342	178	128
	1.2	Sum of weights before poststratification	6,502,472	1,356,078	209,756	458,120	275,221
	1.3	Sum of weights after poststratification	5,924,544	1,268,332	258,400	416,290	288,112
2.	Composite weight						
	2.1	Sum of weights after composite factor	3,413,773	849,821	200,403	280,635	143,340
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	48	6	5	4	2
	3.2	Sum of Weights Before Trimming Adjustment	3,413,773	849,821	200,403	280,635	143,340
	3.3	Sum of Weights After Trimming Adjustment	2,976,850	741,744	195,386	224,628	141,629
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	3,379	659	342	178	128
	4.2	Sum of Weights After Adjustment	3,416,665	990,027	277,643	267,837	149,873
	4.3	Coefficient of Variation (CV)	129.28	102.13	75.77	133.94	122.74
	4.4	Mean Adjustment Factor	1.15	1.34	1.42	1.19	1.06
	4.5	Mean Weight	1011.15	1502.32	811.82	1504.70	1170.88

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum (continued)

			San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	112	130	105	79	63
	1.2	Sum of weights before poststratification	468,242	457,302	334,807	219,140	211,350
	1.3	Sum of weights after poststratification	334,559	419,193	309,644	206,823	194,069
2.	Composite weight						
	2.1	Sum of weights after composite factor	293,566	200,281	127,748	99,181	89,437
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	2	0	0	1	0
	3.2	Sum of Weights Before Trimming Adjustment	293,566	200,281	127,748	99,181	89,437
	3.3	Sum of Weights After Trimming Adjustment	174,025	200,281	127,748	95,274	89,437
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	112	130	105	79	63
	4.2	Sum of Weights After Adjustment	209,633	215,059	119,453	130,337	90,028
	4.3	Coefficient of Variation (CV)	119.27	99.97	130.25	79.90	122.58
	4.4	Mean Adjustment Factor	1.21	1.07	0.94	1.37	1.01
	4.5	Mean Weight	1871.72	1654.30	1137.64	1649.84	1429.02

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum (continued)

		Fresno	San Francisco	Ventura	San Mateo	Kern
1.	Poststratification to telephone service					
1.1	Number of Completed Interviews	71	26	78	36	52
1.2	Sum of weights before poststratification	178,538	196,408	155,989	277,410	202,467
1.3	Sum of weights after poststratification	162,683	174,813	154,918	225,958	132,905
2.	Composite weight					
2.1	Sum of weights after composite factor	80,573	46,322	75,558	99,886	128,088
3.	Trimming Adjustment*					
3.1	Number of Trimmed Records	0	2	0	2	1
3.2	Sum of Weights Before Trimming Adjustment	80,573	46,322	75,558	99,886	128,088
3.3	Sum of Weights After Trimming Adjustment	80,573	45,091	75,558	72,069	68,286
4.	Raking Adjustment*					
4.1	Number of Completed Interviews	71	26	78	36	52
4.2	Sum of Weights After Adjustment	90,911	36,160	72,449	47,177	82,216
4.3	Coefficient of Variation (CV)	82.79	90.74	82.00	89.34	94.16
4.4	Mean Adjustment Factor	1.13	0.80	0.96	0.66	1.20
4.5	Mean Weight	1280.43	1390.76	928.83	1310.48	1581.07

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum (continued)

			San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	50	45	42	46	49
	1.2	Sum of weights before poststratification	173,527	60,354	129,325	62,125	115,246
	1.3	Sum of weights after poststratification	158,224	65,643	121,727	64,703	105,184
2.	Composite weight						
	2.1	Sum of weights after composite factor	58,836	38,555	51,126	33,564	53,177
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	0	0	1	0
	3.2	Sum of Weights Before Trimming Adjustment	58,836	38,555	51,126	33,564	53,177
	3.3	Sum of Weights After Trimming Adjustment	58,836	38,555	51,126	33,158	53,177
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	50	45	42	46	49
	4.2	Sum of Weights After Adjustment	70,144	38,813	57,013	35,182	37,194
	4.3	Coefficient of Variation (CV)	98.86	116.19	71.42	74.67	109.53
	4.4	Mean Adjustment Factor	1.19	1.01	1.12	1.06	0.70
	4.5	Mean Weight	1402.89	862.51	1357.46	764.82	759.06

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum (continued)

			Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	52	37	125	26	37
	1.2	Sum of weights before poststratification	128,981	28,889	15,333	30,964	94,300
	1.3	Sum of weights after poststratification	97,220	30,423	22,742	34,510	87,838
2.	Composite weight						
	2.1	Sum of weights after composite factor	69,006	17,864	20,529	21,304	28,471
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	5	2	1	0	0
	3.2	Sum of Weights Before Trimming Adjustment	69,006	17,864	20,529	21,304	28,471
	3.3	Sum of Weights After Trimming Adjustment	56,865	16,770	19,945	21,304	28,471
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	52	37	125	26	37
	4.2	Sum of Weights After Adjustment	46,689	15,421	15,146	21,067	31,641
	4.3	Coefficient of Variation (CV)	101.38	81.11	96.53	46.15	64.48
	4.4	Mean Adjustment Factor	0.82	0.92	0.76	0.99	1.11
	4.5	Mean Weight	897.86	416.79	121.17	810.26	855.16

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum (continued)

			Merced	Butte	Shasta	Yolo	El Dorado
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	51	44	29	69	33
	1.2	Sum of weights before poststratification	55,826	60,452	13,601	44,373	10,371
	1.3	Sum of weights after poststratification	46,507	44,293	15,283	39,416	15,314
2.	Composite weight						
	2.1	Sum of weights after composite factor	30,685	26,645	9,873	23,824	13,831
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	1	2	0	1	0
	3.2	Sum of Weights Before Trimming Adjustment	30,685	26,645	9,873	23,824	13,831
	3.3	Sum of Weights After Trimming Adjustment	23,672	17,479	9,873	22,657	13,831
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	51	44	29	69	33
	4.2	Sum of Weights After Adjustment	26,308	17,378	12,887	16,403	15,773
	4.3	Coefficient of Variation (CV)	86.96	74.89	71.73	111.89	58.75
	4.4	Mean Adjustment Factor	1.11	0.99	1.31	0.72	1.14
	4.5	Mean Weight	515.84	394.95	444.38	237.73	477.96

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum (continued)

		Imperial	Napa	Kings	Madera	Monterey
1.	Poststratification to telephone service					
1.1	Number of Completed Interviews	52	23	42	44	49
1.2	Sum of weights before poststratification	116,792	18,847	30,135	18,301	83,746
1.3	Sum of weights after poststratification	99,978	20,165	22,338	19,028	69,539
2.	Composite weight					
2.1	Sum of weights after composite factor	20,918	11,447	21,206	10,423	41,939
3.	Trimming Adjustment*					
3.1	Number of Trimmed Records	2	0	2	0	0
3.2	Sum of Weights Before Trimming Adjustment	20,918	11,447	21,206	10,423	41,939
3.3	Sum of Weights After Trimming Adjustment	13,037	11,447	12,819	10,423	41,939
4.	Raking Adjustment*					
4.1	Number of Completed Interviews	52	23	42	44	49
4.2	Sum of Weights After Adjustment	16,182	13,062	11,977	13,552	38,396
4.3	Coefficient of Variation (CV)	95.90	63.32	65.33	75.82	101.79
4.4	Mean Adjustment Factor	1.24	1.14	0.93	1.30	0.92
4.5	Mean Weight	311.19	567.93	285.17	307.99	783.59

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum (continued)

			Humboldt	Nevada	Mendocino	Sutter	Yuba
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	48	31	37	35	44
	1.2	Sum of weights before poststratification	15,725	10,928	41,033	17,741	19,498
	1.3	Sum of weights after poststratification	15,916	10,360	36,300	15,107	16,310
2.	Composite weight						
	2.1	Sum of weights after composite factor	7,553	7,963	8,818	8,292	9,854
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	1	0	0	1	0
	3.2	Sum of Weights Before Trimming Adjustment	7,553	7,963	8,818	8,292	9,854
	3.3	Sum of Weights After Trimming Adjustment	7,071	7,963	8,818	6,529	9,854
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	48	31	37	35	44
	4.2	Sum of Weights After Adjustment	9,031	7,071	7,809	9,391	9,040
	4.3	Coefficient of Variation (CV)	58.97	89.43	83.74	67.14	109.61
	4.4	Mean Adjustment Factor	1.28	0.89	0.89	1.44	0.92
	4.5	Mean Weight	188.15	228.10	211.06	268.31	205.45

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.

Table B-7. Poststratification, composite weighting procedures, trimming and raking adjustments for adolescent interviews by self-reported stratum (continued)

			Lake	San Benito	Tehama, Etc.	Del Norte, Etc.	Tuolumne Calaveras, Etc.
1.	Poststratification to telephone service						
	1.1	Number of Completed Interviews	41	55	35	24	25
	1.2	Sum of weights before poststratification	4,847	4,028	12,956	16,997	66,403
	1.3	Sum of weights after poststratification	6,529	5,452	13,867	16,209	61,723
2.	Composite weight						
	2.1	Sum of weights after composite factor	5,967	4,980	12,009	9,210	21,067
3.	Trimming Adjustment*						
	3.1	Number of Trimmed Records	0	1	0	1	2
	3.2	Sum of Weights Before Trimming Adjustment	5,967	4,980	12,009	9,210	21,067
	3.3	Sum of Weights After Trimming Adjustment	5,967	4,861	12,009	8,700	17,965
4.	Raking Adjustment*						
	4.1	Number of Completed Interviews	41	55	35	24	25
	4.2	Sum of Weights After Adjustment	4,731	5,826	12,086	8,817	13,831
	4.3	Coefficient of Variation (CV)	80.15	86.65	64.97	73.67	64.57
	4.4	Mean Adjustment Factor	0.79	1.20	1.01	1.01	0.77
	4.5	Mean Weight	115.39	105.93	345.31	367.39	553.25

*Counts of completed interviews and sums of weights in sections 1 and 2 are based on self-reported strata rather than sampling strata.