June 2016

CHIS 2013-2014 Methodology Report Series

Report 5

Weighting and Variance Estimation

CALIFORNIA HEALTH INTERVIEW SURVEY

<u>CHIS 2013-2014 METHODOLOGY SERIES</u>

REPORT 5

WEIGHTING AND VARIANCE ESTIMATION

JUNE 21, 2016

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

www.chis.ucla.edu

This report describes the weighting and variance estimation methods used in CHIS 2013-2014. 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 2013-2014 Methodology Series: Report 5 – Weighting and Variance Estimation*. Los Angeles, CA: UCLA Center for Health Policy Research, 2016.

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 2013-2014 came from multiple sources: the California Department of Health Care Services, the California Department of Health Care Services (Mental Health Services Division), the California Department of Public Health, The California Endowment, the California Health Benefit Exchange, the California Health Care Foundation, the California Tobacco Control Program and UC San Diego, the California Wellness Foundation, First 5 California, Kaiser Permanente, Keiro Senior HealthCare, the RAND Corporation, San Diego County Health and Human Services Agency, Sonoma County Department of Health Services, the US Centers for Disease Control and Prevention and the Public Health Institute, and the Williams Institute at UCLA Law School.

PREFACE

Weighting and Variance Estimation is the fifth in a series of methodological reports describing the 2013-2014 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 2013-2014 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 for CHIS 2013-2014

The first five methodological reports for CHIS 2013-2014 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. After the Preface, each report includes an “Overview” chapter (Chapter 1) that is nearly identical across reports, followed by detailed technical documentation on the specific topic of the report.

Report 5: Weighting and Variance Estimation (this report) describes the weighting and variance estimation methods from CHIS 2013-2014. 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)

TABLE OF CONTENTS (CONTINUED)

List of Tables

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

List of Figures

1. CHIS 2013-2014 SAMPLE DESIGN AND METHODOLOGY SUMMARY

This chapter provides a high-level summary of major design components of the California Health Interview Survey (CHIS) and appears at the beginning of each of the five detailed methodology reports for the cycle. You may need to reference those reports to find the level of detail you need. CHIS methodology reports and other methodological documentation and research are online at http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx.

1.1. Overview

This chapter provides a high-level summary of major design components of the California Health Interview Survey (CHIS) and appears at the beginning of each of the five detailed methodology reports for the cycle. You may need to reference those reports to find the level of detail you need. CHIS methodology reports and other methodological documentation and research is online at http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx.

The CHIS is a population-based telephone survey of California's population conducted every other year since 2001 and continually beginning in 2011. CHIS is the largest state health survey and one of the largest health surveys in the nation. CHIS is conducted by the UCLA Center for Health Policy Research (UCLA-CHPR) in collaboration with the California Department of Public Health and the Department of Health Care Services. 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:

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

2) 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. 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. These data are 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. Many researchers throughout California and the nation use CHIS data files to further their understanding of a wide range of health-related issues (for many examples of these studies, visit the Center's publication page (http://healthpolicy.ucla.edu/publications/Pages/default.aspx).

This series of reports describes the methods used in collecting data for CHIS 2013-2014, the sixth CHIS data collection cycle. The previous CHIS cycles (2001, 2003, 2005, 2007, 2009, and 2011-2012) are described in similar series at http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx.

1.2. Switch to a Continuous Survey

From the first CHIS cycle in 2001 through 2009, CHIS data collection was biennial, with data collected during a 7-9 month period every other year. Beginning in 2011, CHIS data have been collected continually over each 2-year cycle. This change was driven by several factors including the ability to track and release information about health in California on a more frequent and timely basis and to eliminate potential seasonality in the biennial data.

CHIS 2013-2014 data were collected between February 2013 and early January 2015. Approximately half of the interviews were conducted during the 2013 calendar year and half during the 2014 calendar year. As in previous CHIS cycles, weights are included with the data files and are based on the State of California's Department of Finance population estimates and projections, adjusted to remove the population living in group quarters (such as nursing homes, prisons, etc.) and thus not eligible to participate in CHIS. When the weights are applied to the data, the results represent California's residential population during that two year period for the age group corresponding to the data file in use (adult, adolescent, or child).

See what's new in the 2013-2014 CHIS sampling and data collection here:
http://healthpolicy.ucla.edu/chis/design/Documents/whats-new-chis-2013-2014.pdf

In order to provide CHIS data users with more complete and up-to-date information to facilitate analyses of CHIS data, additional information on how to use the CHIS sampling weights, including sample code, is available at: http://healthpolicy.ucla.edu/chis/analyze/Pages/sample-code.aspx

Additional documentation on constructing the CHIS sampling weights is available in CHIS 2013-2014 Methods Report #5—Weighting and Variance Estimation, available at:

http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx. Other helpful information for understanding the CHIS sample design and data collection processing can be found in the four other methodology reports for each CHIS cycle year, described in the Preface to this report above.

1.3. Sample Design Objectives

The CHIS 2013-2014 sample was designed to meet the two sampling objectives discussed above: (1) provide estimates for adults in most counties and in groups of counties with small populations; and (2) provide estimates for California's overall population, major racial and ethnic groups, and for several smaller ethnic subgroups.

To achieve these objectives, CHIS employed a dual-frame, multi-stage sample design. The random-digit-dial (RDD) sample included telephone numbers assigned to both landline and cellular service. The random-digit-dial (RDD) sample was designed to achieve completed adult interviews with approximately 80% landline and 20% cellular phone numbers. For the landline RDD sample, the 58 counties in the state were grouped into 44 geographic sampling strata, and 14 sub-strata were created within the two most populous counties in the state (Los Angeles and San Diego). The Los Angeles County stratum included 8 sub-strata for Service Planning Areas, and the San Diego County stratum included 6 sub-strata for Health Service Districts. Most of the strata (39 of 44) are made up of a single county with no sub-strata (counties 3-41 in Table 1-1), with three multi-county strata comprised of the 17 remaining counties (see Table 1-1). CHIS 2013-2014 also included supplemental geographic oversamples of landlines in 3 small counties (Calaveras, Siskiyou, and Tuolumne) that were part of multi-county strata. An address-based sample of an additional 500 households was conducted in Sonoma County and oversamples of about 130 Japanese Americans, 104 Korean Americans, and 120 Vietnamese Americans were completed using list samples. A sufficient number of adult interviews were allocated to each stratum and sub-stratum to support the first sample design objective—to provide health estimates for adults at the local level. The same landline geographic stratification of the state has been used since CHIS 2005. In the first two CHIS cycles (2001 and 2003) there were 47 total sampling strata, including 33 individual counties and one county with sub-strata (Los Angeles).

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.

The RDD CHIS sample is of sufficient size to accomplish the second objective (produce estimates for the state's major racial/ethnic groups, as well as many ethnic subgroups). 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. Surname and given name lists were used similarly to increase the yield of Californians of Japanese descent.

Table 1-1. California county and county group strata used in the CHIS 2013-2014 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, 2013-2014 California Health Interview Survey.

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 2013-2014, the goal was to complete approximately 8,000 interviews (20% of all RDD interviews statewide) with adults from the cell phone sample. Although the geographic information available for cell phone numbers is not as precise as that for landlines, cell phone numbers were assigned to the same 43 strata (i.e., 40 strata defined by a single county and 3 strata created by multiple counties). The cell phone stratification closely resembles that of the landline sample and has the same stratum names, though the cell phone strata represent slightly different geographic areas than the landline strata. As in CHIS 2011-2012, if a sampled cell number was shared by two or more adult members of a household, one household member was selected for the adult interview; otherwise the adult owner of the sampled number was selected. Cell numbers used exclusively by children under 18 were considered ineligible. About 480 teen interviews and 1,250 child interviews were completed from the cell phone sample in CHIS 2013-2014.

The cell phone sampling method used in CHIS has evolved since its first implementation in 2007 when only cell numbers belonging to adults in cell-only households were eligible for sampling adults. There have been two significant changes to the cell phone sample since 2009. First, all cell phone sample numbers used for non-business purposes by adults living in California were eligible for the extended interview. Thus, adults in households with landlines who had their own cell phones or shared one with another adult household member could have been selected through either the cell or landline sample. The second change was the inclusion of child and adolescent extended interviews.

The cell phone sample design and targets by stratum of the cell phone sample have also changed throughout the cycles of the survey. In CHIS 2007 a non-overlapping dual-frame design was implemented where cell phone only users were screened and interviewed in the cell phone sample. Beginning in 2009, an overlapping dual-frame design has been implemented. In this design, dual phone users (e.g., those with both cell and landline service) can be selected and interviewed from either the landline or cellphone samples.

The number of strata has also evolved as more information about cell numbers has become available. In CHIS 2007 the cell phone frame was stratified into 7 geographic sampling strata created using telephone area codes. In CHIS 2009 and 2011-2012, the number of strata was increased to 28. These strata were created using both area codes and the geographic information assigned to the number. In CHIS 2011-2013, with the availability of more detailed geographic information, the number of strata was increased to 43 geographic areas that correspond to single and grouped counties similar to the landline strata.

1.4. Data Collection

To capture the rich diversity of the California population, interviews were conducted in six languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, Korean, and, for the first time, Tagalog. These languages were chosen based on analysis of 2010 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 CHIS 2013-2014 data collection under contract with the UCLA Center for Health Policy Research. For all samples, Westat staff interviewed one randomly selected adult in each sampled household, and sampled one adolescent and one child if they were present in the household and the sampled adult was their parent or legal guardian. Thus, up to three interviews could have been completed in each household. Children and adolescents were generally sampled at the end of the adult interview. In landline, list, and ABS sample households with children where the screener respondent was someone other than the sampled adult, 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 first used in CHIS 2005 and has been continued in subsequent CHIS cycles because it substantially increases the yield of child interviews. While numerous subsequent attempts were made to complete the adult interview for child-first cases, the final data contain completed child and 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 2013-2014 by the type of sample (landline RDD, surname list, cell RDD, and Sonoma ABS). These numbers are provided in greater detail in Chapter 6 of this report/ *CHIS 2013-2014 Methodology Series: Report 2 – Data Collection*. Note that these figures were accurate as of data collection completion and may differ slightly from numbers in the data files due to data cleaning and edits. Sample sizes to compare against data files you are using are found online at http://healthpolicy.ucla.edu/chis/design/Pages/sample.aspx.

Table 1-2. Number of completed CHIS 2013-2014 interviews[1] by type of sample and instrument

Type of sample	Adult*	Child	Adolescent
Total all samples	40,240	5,512	2,253
Landline RDD	31,615	4,164	1,738
Surname list	392	50	18
Cell RDD	7,752	1,256	482
Sonoma ABS	481	42	15

*Includes interviews meeting the criteria as partially complete

Source: UCLA Center for Health Policy Research, 2013-2014 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 36 minutes to complete. The average

[1] Numbers in this table represent the data publically released and available through our Data Access Center. Total sample sizes may differ for specific calculations within the five methodology reports, or for specific analyses based on CHIS data.

child and adolescent interviews took about 16 minutes and 23 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 11 percent of the adult interviews were completed in a language other than English, as were about 23 percent of all child (parent proxy) interviews and 5 percent of all adolescent interviews.

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

Table 1-3. CHIS 2013-2014 survey topic areas by instrument

Health status	**Adult**	**Teen**	**Child**
General health status	✓	✓	✓
Days missed from school due to health problems		✓	✓
Health conditions	**Adult**	**Teen**	**Child**
Asthma	✓	✓	✓
Diabetes, gestational diabetes, pre- /borderline diabetes	✓		
Heart disease, high blood pressure	✓		
Physical disability	✓		
Physical, behavioral, and/or mental conditions			✓
Mental health	**Adult**	**Teen**	**Child**
Mental health status	✓	✓	
Perceived need, access and utilization of mental health services	✓	✓	
Functional impairment, stigma	✓		
Suicide ideation and attempts	✓	✓	
Health behaviors	**Adult**	**Teen**	**Child**
Dietary intake, fast food	✓	✓	✓
Physical activity and exercise, commute from school to home		✓	✓
Walking for transportation and leisure	✓		
Doctor discussed nutrition/physical activity		✓	
Flu Shot	✓	✓	✓
Cigarette use, second-hand smoke, attitudes about smoking	✓	✓	
Alcohol use	✓	✓	
Sexual behavior	✓		
HIV/STI testing	✓		
Sedentary time		✓	✓
Dental health	**Adult**	**Teen**	**Child**
Last dental visit	✓	✓	✓
Main reason haven't visited dentist	✓	✓	
Current dental insurance coverage	✓	✓	✓

Table 1-3. CHIS 2013-2014 survey topic areas by instrument (continued)

Neighborhood and housing	**Adult**	**Teen**	**Child**
Social cohesion	✓	✓	✓
Neighborhood safety	✓	✓	✓
Homeownership, length of time at current residence	✓		
Park use		✓	✓
Civic engagement	✓	✓	
Access to and use of health care	**Adult**	**Teen**	**Child**
Usual source of care, visits to medical doctor	✓	✓	✓
Emergency room visits	✓	✓	✓
Inpatient hospital stays	✓		
Delays in getting care (prescriptions and medical care)	✓	✓	✓
Patient-centered care, timely appointments, care coordination	✓	✓	✓
Communication problems with doctor	✓		✓
Problems finding a doctor	✓		✓
Use of specialists	✓		
Advance directive (Sonoma County)	✓		
Internet use for health information	✓		✓
Contraception (counseling, prescription, male birth control)	✓		
Food environment	**Adult**	**Teen**	**Child**
Access to fresh and affordable foods	✓		
Fast food at school, School lunch consumption		✓	✓
Water availability		✓	
Water consumption	✓	✓	
Availability of food in household over past 12 months	✓		
Health insurance	**Adult**	**Teen**	**Child**
Current insurance coverage, spouse's coverage, who pays for coverage	✓	✓	✓
Health plan enrollment, characteristics and plan assessment	✓	✓	✓
Whether employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months, reasons for lack of insurance	✓	✓	✓
Coverage through Covered California	✓	✓	✓
Difficulty finding private health insurance	✓	✓	✓
High deductible health plans	✓	✓	✓
Partial scope Medi-Cal	✓		
Public program eligibility	**Adult**	**Teen**	**Child**
Household poverty level	✓		
Program participation (CalWORKs, Food Stamps/CalFresh, SSI, SSDI, WIC, TANF)	✓	✓	✓
Assets, alimony/child support, social security/pension	✓		
Medi-Cal and Healthy Families eligibility	✓	✓	✓
Reason for Medi-Cal non-participation among potential beneficiaries	✓	✓	✓

Table 1-3. CHIS 2013-2014 survey topic areas by instrument (continued)

Parental involvement/adult supervision	**Adult**	**Teen**	**Child**
Parental involvement			✓
Child care and school attendance	**Adult**	**Teen**	**Child**
Current child care arrangements			✓
Paid child care	✓		
First 5 California: Kit for New Parents			✓
Preschool/school attendance, name of school		✓	✓
Preschool quality			✓
Special programs in school		✓	
Grades, college expectations		✓	
Organizational involvement, civic engagement		✓	
School instability		✓	
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, number of persons supported by household income	✓		
Alimony/child support	✓		
Worker's compensation, Social Security, pensions	✓		
Respondent characteristics	**Adult**	**Teen**	**Child**
Race and ethnicity, age, gender, height, weight	✓	✓	✓
Veteran status	✓		
Marital status, registered domestic partner status (same-sex couples)	✓		
Sexual orientation	✓		
Language spoken with peers, language of TV, radio, newspaper used	✓		
Education, English language proficiency	✓		
Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home	✓	✓	✓

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

1.5. Response Rates

The overall response rate for CHIS 2013-2014 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 50.7 percent of the landline RDD sample telephone numbers not identified by the sample vendor as business or nonworking numbers, and for 82.2 percent of surname list sample numbers. Addresses were not available for the cell sample. As in all CHIS cycles since CHIS 2005, a $2 bill was included with the CHIS 2013-2014 advance letter to encourage cooperation.

The CHIS 2013-2014 screener response rate for the landline/list sample was 28.8 percent, and was higher for households that were sent the advance letter. For the cell phone sample, the screener response rate was 30.7 percent. The extended interview response rate for the landline/list sample varied across the adult (44.8 percent), child (68.9 percent) and adolescent (40.2 percent) interviews. The adolescent rate includes getting permission from a parent or guardian. The adult interview response rate for the cell sample was 52.1 percent, the child rate was 72.2 percent, and the adolescent rate 41.0 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 2013-2014, the landline/list sample household response rate was 14.8 percent (the product of the screener response rate and the extended interview response rate at the household level of 51.4 percent). The cell sample household response rate was 16.6 percent, incorporating a household-level extended interview response rate of 53.9 percent. All of the household and person level response rates vary by sampling stratum. For more information about the CHIS 2013-2014 response rates please see *CHIS 2013-2014 Methodology Series: Report 4 – Response Rates*.

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. However, comparing the CHIS and BRFSS response rates requires recomputing the CHIS response rates so they match the BRFSS response rate calculation methods. The 2013 California BRFSS landline response rate is 38.9 percent, the cell phone response rate is 39.3 percent, and the combined landline and cell phone rate is 39.0 percent.[2] Recalculating the CHIS response rates using the BRFSS method, the CHIS 2013-2014 landline response rate is 39.5, cell phone response rate is 32.1 percent, and the combined landline and cell phone response rate is 37.2 percent. 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. For example, based on the last reported BRFSS refusal rates in 2011; the

[2] As reported in the Behavioral Risk Factor Surveillance System: 2013 Summary Data Quality Report. Retrieved May 22, 2015, available online at http://www.cdc.gov/brfss/annual_data/2013/pdf/2013_dqr.pdf

refusal rate for California (31.4%) was the highest in the nation and was twice the national median (16.0%). 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://healthpolicy.ucla.edu/chis/design/Pages/data-quality.aspx.

After all follow-up attempts to complete the full questionnaire were exhausted, adults who completed at least approximately 80 percent of the questionnaire (i.e., through Section K which covers employment, income, poverty status, and food security), were counted as "complete." At least some responses in the employment and income series, or public program eligibility and food insecurity series were missing from those cases that did not complete the entire interview. They were imputed to enhance the analytic utility of the data (see section 2.6 on imputation methods for more information).

Proxy interviews were conducted for any adult who was unable to complete the extended adult interview for themselves, in order to avoid biases for health estimates of chronically-ill or handicapped people. Eligible selected persons were re-contacted and offered a proxy option. For 248 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.

1.6. Weighting the Sample

To produce population estimates from 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 2013-2014 accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons;
- Reduce biases occurring because non-respondents may have different characteristics than respondents;
- Adjust, to the extent possible, for under-coverage 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 an iterative proportional fitting method, or “raking” as it is commonly called, so that CHIS estimates are consistent with the marginal population control totals. This iterative procedure 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 pre-specified tolerance.

Population control totals of the number of persons by age, race, and sex at the stratum level for CHIS 2013-2014 were created primarily from the California Department of Finance’s (DOF) 2014 Population Estimates and 2014 Population Projections. The raking procedure used 12 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 excluding households without landline telephones from the sample frame. One limitation of using Department of Finance (DOF) 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 and includes, for example nursing homes, prisons, dormitories, etc.). 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.

The 2014 DOF control totals used to create the CHIS 2013-2014 weights are based on 2010 Census counts, as were those used for the 2011-2012 cycle. Please pay close attention when comparing estimates using CHIS 2013-2014 data with estimates using data from CHIS cycles before 2010. The most accurate California population figures are available when the US population count is conducted (every 10 years). Population-based surveys like CHIS must use estimates and projections based on the decennial population count data between Censuses. For example, population control totals for CHIS 2009 were based on 2009 DOF estimates and projections, which were based on Census 2000 counts with adjustments for demographic changes within the state between 2000 and 2009. These estimates become less accurate and more dependent on the models underlying the adjustments over time. Using the most recent Census population count information to create control totals for weighting produces the most statistically accurate population estimates for the current cycle, but it may produce unexpected increases or decreases in some survey estimates when comparing survey cycles that use 2000 Census-based information and 2010 Census-based information. See *CHIS 2013-2014 Methodology Series: Report 5 – Weighting and Variance Estimation* for more information on the weighting process.

1.7. 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 every other variable.

Two different imputation procedures were used by Westat to fill in missing responses 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 one of the most commonly used methods 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 who answer a survey item form a pool of donors, while the item non-respondents form 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. Westat used hot deck imputation to impute the same items in all CHIS cycles since 2003 (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 2013-2014 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 20%, 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 corresponds to the nature of the variable being imputed (e.g. linear regression for continues variables, logistic regression for binary variables, etc.). 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 standard measures of demographic and socioeconomic characteristics, as well as geographic region; however, the full set of control variables varies depending on which variable is being imputed. Most imputation models included additional characteristics, such as health status or access to care, which are used to improve the quality of the donor-recipient match. Among the standard list of control variables, gender, age, race/ethnicity and region of California were imputed by Westat. UCLA-CHPR began their imputation process by imputing household income and educational attainment, so that these characteristics were available for the imputation of other variables. CHIS collects bracketed information about the range in which the respondent's value falls when the respondent will not or cannot report an exact amount. Household income, for example, was imputed using the hot-deck method within ranges defined by a set of auxiliary variables such as bracketed income range and/or poverty level. After all other variables are imputed, household income is re-imputed using a more detailed list of covariates to create a higher quality match between donors and recipients.

The imputation order of the other variables generally followed their order in 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 non-imputed values on a case-by-case basis.

1.8. Methodology Report Series

A series of five methodology reports is available with more detail about the methods used in CHIS 2013-2014:

- 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

This chapter introduces the concept of weighting and provides some background on the weights developed for analyzing CHIS 2013-2014 survey data. Weighting is a post-data collection process that supports estimates from data provided by survey respondents representative of the total population from which they were sampled. Weighting accounts for the chances of selecting units into the sample and adjusts for imperfections in the frame and data collection. The process begins with a base weight that is adjusted to account for additional stages of sampling, nonresponse and undercoverage.

As described in *CHIS 2013-2014 Methodology Series: Report 1 - Sample Design*, CHIS 2013-2014 has samples from four different sampling frames; landline, cell phone, supplemental surname lists, and an addressed-based sample (ABS) frame. One set of weights was produced for data analysis for the combined 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 2013-2014 weighting is a standard design-based, multiple-frame methodology that is consistent with the sampling methods used during sample selection. The multiple-frame approach has been used since CHIS 2009 to combine and weight the landline, cell, and list telephone samples. In CHIS 2013-2014, the same approach was used to weight the Sonoma address-based sample (ABS) and to combine the ABS sample with the landline, cell, and list samples of telephone numbers.

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 supplemental samples. 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.

Weights are applied to CHIS 2013-2014 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 telephone numbers, addresses, and associated 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.

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), the cell phone base weights (Section 3.1.2), and ABS base weights (Section 3.1.3), the base weights are adjusted for the following when applicable:

- Remaining ported cell numbers not dialed during data collection (Section 3.2);
- Sampled telephone numbers that were 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 respondents (Section 3.8).

These adjustments are described in Chapter 3.

The final 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);
- Trimming (Section 4.4) and raking (Section 4.5) adjustment to person-level control totals; and
- Composite weight adjustment and raking to combine landline/cell phone/list weights and ABS weights (Section 4.6).

The child and adolescent weights are more complex because of the method used to sample children (see CHIS 2013-2014 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 (Sections 5.2 and 6.1);
- Extended child and adolescent interview nonresponse adjustment (Sections 5.3 and 6.2);
- Telephone type adjustment (Sections 5.3 and 6.2);
- Composite weight adjustment for combining the landline and cell phone samples (Sections 5.3 and 6.2);
- Trimming and raking (Sections 5.3 and 6.2) adjustment to person-level control totals; and
- Composite weight adjustment and raking for combining of landline/cell phone/list weights and ABS weights (Section 5.3 and 6.2).

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 2013-2014.

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 2013-2014, the weights of those who respond are adjusted to represent noncovered 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 (Groves, 1989). The purpose of adjusting for nonresponse is to reduce the bias. A weighting class adjustment method (Brick & Kalton, 1996) is the type of nonresponse adjustment procedure used in CHIS 2013-2014. 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 2013-2014 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 & 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 or address), 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.

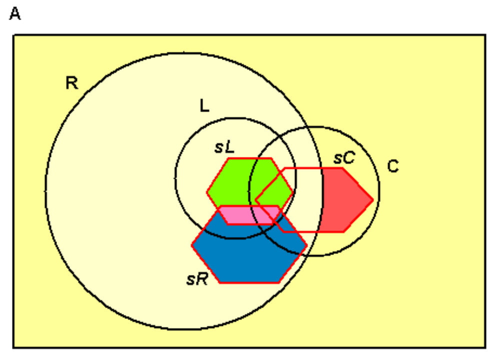
The approach to adjusting for undercoverage is somewhat different from that for nonresponse because noncovered units or persons were never eligible to be sampled. The undercoverage adjustment procedure uses data from external sources (control totals) in a process called poststratification (Holt & Smith, 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 2013-2014 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 (Brackstone & Rao, 1979). In CHIS 2013-2014, the control totals are primarily derived from the 2014 California Department of Finance Population Estimates and Projections (State of California, Department of Finance, 2013, 2014), the 3-year 2011-2013 American Community Survey (U.S. Census Bureau, 2014), the Census 2010 Summary File 1 for California published by the U.S. Census Bureau (U.S. Census Bureau, 2012), and the Census 2010 Modified Race Data (U.S. Census Bureau, 2012). Creation of the control totals at the person level is described in Chapter 7.

2.4. Combining the Landline, Cell phone, and List Samples

In this section, we describe how the landline, cell phone, and list samples were combined to create the weights for CHIS 2013-2014. Before explaining the approach for combining the samples, we examine the relationship between the landline and cell phone frames and samples.

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

Figure 2-1. Landline, list, and cell phone frames and samples in CHIS 2013-2014



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 (LA) County (represented by the large yellow rectangle in the diagram). Let *R* (the large circle in the diagram) be all LA County households with a telephone number in the landline frame, and *L* (smaller circle enclosed within R) be all LA County households with a telephone number in the surname list frame. Note that by definition, *R* is included within *A* and that *L* is included within *R* (i.e., $A \subset R \subset L$). 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., $C \cap \bar{R}$), 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 list or surname frame have two probabilities of selection (one from the landline sample and the other from 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 list samples. Thus, the landline and list 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. 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 one or more cell phones only (i.e., $C \cap \bar{R}$) and households with both telephone services (i.e., $C \cap R$) were eligible in CHIS 2013-2014. Their base weights were computed as the inverse of the probability of selection from the respective frames. 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. 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

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

where λ ($0 \le \lambda \le 1$) is the composite or weighting factor. In CHIS 2013-2014, the value of λ was chosen to minimize the bias of $\hat{Y}$. The choice is outlined in Brick, Flores Cervantes, Lee, & Norman, (2011) and differs from the Hartley 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 2013-2014, a composite weight was created rather than requiring that calculation of every estimate from CHIS include the composite factor. 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

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

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 surname 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.

2.5. Combining the ABS and Landline/Cell phone/List Samples

For the first time in CHIS, the 2013-2014 samples included a supplemental ABS sample used to increase the number of interviews in Sonoma County (additional details of the Sonoma ABS sample can be found in *CHIS 2013-2014 Methodology Series: Report 1 – Sample Design)*. Using a procedure similar to that used to combine the landline and cell phone samples described in the previous section, a single weight that combined the ABS sample with the combined landline/cell phone/ list weight was created in CHIS 2013-2014. Since the ABS sample and the combined landline/cell phone/ list sample for Sonoma County were drawn independently and were each representative of the population in the county, a combined weight with a composite factor ϕ can be used to produce estimates that combine the samples as

$$\hat{Y} = \sum_{i \in ABS} w_i \phi y_i + \sum_{i \in L/C/L} w_i \left(1-\phi\right) y_i \ .$$

The samples are combined at the last of step of the weighting process. Additional details of the creation of the composite weight that combined these samples are found in Section 4.6.

3. HOUSEHOLD WEIGHTING

For all CHIS 2013-2014 samples, the first step in the weighting process 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 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. Some of these adjustments are applicable to specific samples. Adjustments to the landline sample account for 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. Adjustments to the cell phone sample weights account for subsampling of cell phone numbers based on the activity flag, unknown residential status, and screener nonresponse. Some of these adjustments are not applicable to the ABS sample.

Knowledge of the sampling methods used in CHIS 2013-2014 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 2013-2014 Methodology Series: Report 1 - Sample Design*.

3.1. Base Weights

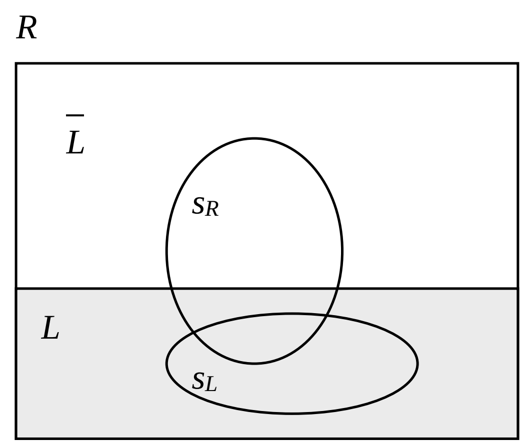
A base weight is created for each sampling unit in the different CHIS samples. For the landline, list, and cell samples, the sampling unit is the telephone number. In contrast, the sampling unit is a mailing address in the ABS sample

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 2013-2014, telephone numbers were drawn from the landline frame, six mutually exclusive surname frames (Korean only surname, Vietnamese only surname, Korean and any other race/ethnic surname but Vietnamese surname, Japanese first names, Japanese last names, and Japanese first and last names). The base weights reflect the multiple probabilities of selection of telephone numbers between the landline and different list frames.

Figure 3-1 shows the relationship between the landline frame and a single 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.

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

The notation in the figure follows:

R	the landline frame containing all telephone numbers assigned to landline service;
L	the list frame (i.e., surnames or clinic users, and associated landline telephone numbers);
$\bar{L}$	all landline 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 the 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

$$N_{Rh} = 100 \cdot \frac{NBANKS_{2013,h} + NBANKS_{2014,h}}{2},$$

where $NBANKS_{2013,h}$ and $NBANKS_{2014,h}$ are the number of 1+ banks in the stratum h in the 2013 and 2014 landline frames respectively. A “1+” bank is defined as a 100-bank with at least one listed 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 $\left(NL_{Lh}\right)$ is the number of records in the list assigned to stratum h.

As described in *CHIS 2013-2014 Methodology Series: Report 1 - Sample Design*, the landline sample was drawn from strata defined as counties or groups of counties except for Los Angeles County, San Diego County, Orange County, and Santa Clara County. In Los Angeles County, 13 subsampling strata were created by the combination of areas with high/low concentration of Koreans and Vietnamese and eight Service Planning Areas (SPAs). In San Diego County, eight subsampling strata were created by the combination of areas with high/low concentration of Koreans and Vietnamese and six Health Service Regions (SRs). Two substrata based on the concentration of Koreans and Vietnamese were created for Orange and Santa Clara Counties. The definition of the sampling strata and substrata and the number of telephone numbers in the landline frame, the number of sample cases, and base weights by frame type (landline, Korean only surname, Vietnamese only surname, Korean and another group but not Vietnamese surname, Japanese first name, Japanese last names, and Japanese first and last names), are shown in Appendix A, Table A-1. 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] A bank is defined as 100 consecutive telephone numbers with the same first eight digits including area code.

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 communication service types 65, 68)[4]. The sampling strata were defined by the area code of telephone numbers assigned to wireless service and pre-assigned FIPS county code. For more details on the cell phone samples, see *CHIS 2013-2014 Methodology Series: Report 1 - Sample Design*.

The cell sample base weight is computed in the same way as the landline/list base weight described in Section 3.1.1. The only difference is that the total number of telephone numbers in the cell phone frame in stratum *h* is computed using 1,000 blocks. Note that the stratum definition for the cell phone sample is different from that of the landline sample in that they do not match the same geographic areas as in the landline sample for most strata, and they do not include separate values for Los Angeles SPAs, San Diego health regions, and for the geographic areas with high/low concentrations of Koreans and Vietnamese in some counties.

The definitions of the sampling strata and substrata and average base weights, are shown in Appendix A, Table 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 the main cell phone sample.

3.1.3. ABS Base Weight

The ABS sample was drawn from a frame of mailing addresses in Sonoma County. The frame was created by identifying the geographic location (i.e., geocode) of addresses in Sonoma County. The base weight for the ABS sample, $HHBW_i$, is computed as

$$HHBW_i = \frac{N_A}{n_A};$$

where N_A is the total number of addresses in the frame and n_A is the total numbers of sampled addresses in Sonoma County. Table B-1 in Appendix B (rows 1.1 through 1.3) shows the sample counts, sums of base weights, and coefficients of variation of the base weights of the ABS sample.

[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.

3.2. Cell Phone Activity Status Adjustment

Due to new agreements between cell phone service providers and sampling vendors, the activity status of a cell phone number can be determined without dialing for most cell phone numbers. The activity status of the cell phone sample was used to increase the efficiency of the sample by dialing only a subsample of cell phone numbers. In CHIS 2013-2014, subsampling based on the activity flag was implemented in the second half of data collection. The cell phone base weights were adjusted to account for those numbers subsampled as inactive or with an unknown activity status. Before the adjustment, the cell phone sample was classified into three groups as indicated in Table 3-1 depending on the value of the activity flag (i.e., active, inactive, or unknown activity status). As noted in the table, those not identified as active were subsampled and the telephone numbers subsampled were retained and dialed and those not subsampled were not dialed).

Table 3-1. Groups based on the cell phone activity

Activity flag group	Description
ACTIVE	Cell phone likely to be active
RET_YES	Cell phone likely to be inactive or with unknown activity status retained in the sample and dialed
RET_NO	Cell phone likely to be inactive or with unknown activity status not retained in the sample

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

The weights of the dialed inactive/unknown cell phone numbers were adjusted to account for those inactive/unknown cell phone numbers that were not retained in the sample. The activity flag telephone adjusted weight, $HHA1W_i$, is computed as

$$HHA1W_i = HHA1F_i * HHBW_i ,$$

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

$$HHA1F_i = \begin{cases} 1 & \text{If } i \in ACTIVE \\ \dfrac{\sum\limits_{i \in RET\_YES, RET\_NO} HHBSW_i}{\sum\limits_{i \in RET\_NO} HHBSW_i} & \text{If } i \in RET\_YES \\ 0 & \text{If } i \in RET\_NO \end{cases},$$

where the groups ACTIVE, RET_YES, and RET_NO are defined in Table 3-1. This adjustment was done separately by sampling stratum. The adjustment was only applied to numbers selected from the cell phone frame so the adjustment factor $HHA1F_i$ was set to one for all records in the other samples. Table B-1 in Appendix B (rows 2.1 through 2.4) shows the sample size, coefficient of variation and sums of weights before and after the adjustment.

3.3. New Work Adjustment

Additional telephone numbers were drawn during data collection depending on the number of completed interviews achieved to date and the projected number of completed interviews at the end of the data collection period. However, not all newly drawn telephone numbers were dialed because the targets in some strata were met before exhausting the sample in those strata. As a result, the weights were adjusted to account for sampled numbers 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, $HHA2W_i$, was 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 sums 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 in households in which a member refused to participate in the study at the screener level; shifting some resources from the less productive, labor-intensive task of refusal conversion to the more productive task of completing extended interviews increased the efficiency of data collection. Due to changes over time in the relative efficiency of different kinds of work, subsampling for refusal conversion was no longer done after 2007, including 2013-2014. Since then, all refusals in the landline and surname samples were eligible for two refusal conversion attempts at the screener level if neither refusal was judged to be hostile or abusive. Starting in CHIS 2011-2012, second refusal conversion was also implemented for the cell sample. 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, the full refusal conversion protocol was not carried out because the data collection period ended. In this adjustment, the weights of the cases with two refusal conversion attempts were adjusted to account for the few cases that have none or only one refusal conversion. This adjustment assumes that refusals without refusal conversion attempts were a random sample of those with refusal conversion attempts.

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 3-2.

Table 3-2. 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, 2013-2014 California Health Interview Survey.

The first refusal subsampling-adjusted weight, $HHA3W_i$, was computed as

$$HHA3W_i = HHA3F_i * HHA2W_i$$

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

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

where the groups $RefC1$, $RefNC$, and $N\,\mathrm{Re}\, f$ are defined in Table 3-2, $HHA2W_i$ 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:

$$HHA4W_i = HHA4F_i * HHA3W_i$$

where $HHA4F_i$ is the second refusal subsampling adjustment factor computed as:

$$HHA4F_i = \begin{cases} \dfrac{\sum\limits_{i \in (RefC2, RefNC1)} HHA3W_i}{\sum\limits_{i \in RefC2} HHA3W_i} & \text{If } i \in RefC2 \\ 0 & \text{If } i \in \mathrm{Re}\, fNC \\ 1 & \text{If } i \in NRef \end{cases},$$

where the groups $RefC2$, $RefNC$, and $N\,\mathrm{Re}\,f$ are defined in Table 3-2. 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 in the landline, cell phone, and landline samples 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 2013-2014 Methodology Series: Report 4 - Response Rates*.

In CHIS 2013-2014, the estimated proportion of unknown residential telephone numbers considered residential $\left(p_{res}\right)$ was computed separately for the landline, surname, and cell phone samples. The proportion p_{res} was computed following the CASRO recommendation (Council of American Survey Research Organizations, 1982) 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 was used rather than the (unweighted) number of cases to compute p_{res}. This use of weights also compensates for the under- and oversampling implemented in different geographic areas.

There are some differences in the way the value of values of p_{res} was computed for the samples. These differences are based on the information available by type of sample. Table 3-3 shows the values of

p_{res} for the landline sample, calculated separately for each combination of mail status, urbanicity, and paradata on the type of household based on the answering machine. The type of household was determined by the interviewer after listening to the greeting in the answering machine or voice mail. 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 numbers not associated with an address that had answering machine messages coded as possible nonresidential or unknown.

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

Mail status	Urban status	Answering machine code	p_{res}
Mailable	Urban	No machine	0.617
Mailable	Urban	Possible residential	0.903
Mailable	Urban	Possible nonresidential	0.351
Mailable	Urban	Unknown	0.869
Mailable	Not urban	No machine	0.691
Mailable	Not urban	Possible residential	0.909
Mailable	Not urban	Possible nonresidential	0.362
Mailable	Not urban	Unknown	0.877
Not mailable	Urban	No machine	0.194
Not mailable	Urban	Possible residential	0.841
Not mailable	Urban	Possible nonresidential	0.140
Not mailable	Urban	Unknown	0.648
Not mailable	Not urban	No machine	0.228
Not mailable	Not urban	Possible residential	0.849
Not mailable	Not urban	Possible nonresidential	0.116
Not mailable	Not urban	Unknown	0.673

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

Since the list samples were relatively small, the values of p_{res} were computed at the state level (e.g., most list samples have a mailing address and are mainly urban). Table 3-4 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 3-4. Estimated residential proportion for the list samples

Answering machine code	p_{res}
No machine	0.392
Answering machine possible residential	0.885
Answering machine possible nonresidential	0.210
Answering machine unknown	0.799

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

In the cell phone sample, there is no mailing address to determine the mailable status of the cell phone and the paradata of the answer machine had not enough variation to be useful for the creation of groups by answering machine code. As a result, the values of p_{res} were computed by sampling stratum in the cell phone sample as shown in Table 3-5.

Table 3-5. Estimated residential proportion for the RDD cell phone samples by sampling strata

Sampling stratum	Counties covered	p_{res}
1	Los Angeles	0.622
2	San Diego	0.609
3	Orange	0.655
4	Santa Clara	0.696
5	San Bernardino	0.607
6	Riverside	0.632
7	Alameda	0.617
8	Sacramento	0.694
9	Contra Costa	0.705
10	Fresno	0.584
11	San Francisco	0.692
12	Ventura	0.658
13	San Mateo	0.678
14	Kern	0.571
15	San Joaquin	0.670
16	Sonoma	0.663
17	Stanislaus	0.640
18	Santa Barbara	0.640
19	Solano	0.734
20	Tulare	0.611
21	Santa Cruz	0.714
22	Marin	0.629
23	San Luis Obispo	0.629
24	Placer	0.472
25	Merced	0.636
26	Butte	0.681
27	Shasta	0.681
28	Yolo	0.702

Table 3-5. Estimated residential proportion for the RDD cell phone samples by sampling strata (continued)

Sampling stratum	Counties covered	$\hat{p}_{res}$
29	EL Dorado	0.636
30	Imperial	0.577
31	Napa	0.697
32	Kings	0.670
33	Madera	0.609
34	Monterey	0.586
35	Humboldt	0.612
36	Nevada	0.705
37	Mendocino	0.591
38	Sutter, Yuba	0.676
40	Lake	0.504
41	San Benito	0.651
42	Colusa, Glen, Tehama	0.512
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	0.503
44	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	0.513

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

The estimated proportion of residential households among the unknown residential telephone numbers, p_{res}, was 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:

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

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

This adjustment was done separately by sample type. In the landline sample, the adjustment was done within sampling stratum by mailable status. In the list sample, the adjustment was by list type. In the cell phone sample, the weights were adjusted by sampling stratum and ported and non-ported number. This adjustment is not applicable to the ABS sample because we assumed that all sampled addresses are residential households. The adjustment factor $HHA3F_i$ was set to one for all records in the ABS sample. Table B-1 in Appendix B (rows 6.1 through 6.5) shows the sums of weights before and after the adjustment for unknown residential status for the landline, surname, and cell samples.

3.6. Sample Eligibility Nonresponse Adjustment

After the unknown residential status adjustment, the weights were adjusted for eligibility as determined in the screener. In CHIS 2013-2014, screening was used only to identify households with adults of Korean, Vietnamese, or Japanese descent in the surname samples. Therefore, this adjustment was 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, Vietnamese, or Japanese) could not be determined. Telephone numbers from the list samples were eligible only if one or more adults in the household considered themselves of Korean, Vietnamese, or Japanese descent.[5]

Households with at least one adult from one of these 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 groups was selected for the extended interview. If the household was not list-eligible (i.e., no adults of Korean, Vietnamese, or Japanese 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 supplemental list samples was a relatively efficient method for increasing the number of Korean, Vietnamese, or Japanese extended interviews in CHIS 2013-2014 and previous cycles. Information on the ethnic origin or race of the adults was used to avoid unnecessary interviews of adults from a different group, who were represented adequately in the landline sample.

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

[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} {Japanese} descent?"

observed proportion in the sample with known eligibility. The cases were classified in response groups indicated in Table 3-6.

Table 3-6. List eligibility response groups

List eligibility response status group		Description
L_E	List-eligible	Household from the list sample with at least one list-eligible adult (i.e., adult of Korean, Vietnamese, or Japanese descent).
L_IN	List-ineligible	Household from the list sample without any list-eligible adult (i.e., no adults of Korean, Vietnamese or Japanese descent).
L_UNK	List eligibility unknown	Household from the 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).

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

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

$$HHA6W_i = HHA6F_c * HHA5W_i ,$$

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

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

where the groups L_E, L_IN, L_UNK, and L_NA are defined in Table 3-6, 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 three Japanese samples). This weighting adjustment is not applicable to the cell phone and ABS samples. Table B-1 in Appendix B (rows 7.1 through 7.4) shows the sums of weights before and after the list eligibility nonresponse adjustment.

3.7. Screener Nonresponse Adjustment

In this step, the household weights in the landline, cell phone, and list samples were 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 unknown child status were distributed to responding households. This weight, $HHA7W_i$, is computed as

$$HHA7W_i = HHA7F_c * HHA6W_i,$$

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

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

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 the indicator $\delta_i(c)$ 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 sums of weights before and after the unknown presence of children in household adjustment. This adjustment is not applicable to the ABS sample because the presence of children in the household was not collected in the mailed screener form.

In the second step of screener nonresponse adjustment we adjusted the weights to account for screener nonresponse. This weight, $HHA8W_i$, is:

$$HHA8W_i = HHA8F_c * HHA7W_i,$$

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

---

[6] There are differences in response rates between households with and without children. See *CHIS 2013-2014 Methodology Series: Report 4–Response Rates.*

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

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, Vietnamese, or Japanese 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) shows the sums of weights before and after the screener nonresponse adjustment.

The form of the screener nonresponse adjustment was different for the ABS sample, and depended on the procedure used to obtain the telephone number for the sampled addresses and the different types of nonresponse during this process. After selection, the sampled addresses were matched to telephone numbers using reverse telephone matching services. The remaining addresses without a matched telephone number were mailed a screening questionnaire asking for a telephone number associated with the address. All available telephone numbers, whether from the matching process or the mail screener, were dialed (see *CHIS 2013-2014 Methodology Series: Report 1 – Sample Design* and *CHIS 2013-2014 Methodology Series: Report 2 - Data Collection Methods* for additional details).

Nonresponse occurred in those households that did not return the mail screener, those households with telephone numbers that were not contacted, and those contacted households that refused the telephone interview. The combination of the disposition codes of the mailed returns and telephone calls was used to classify the ABS sample into the screener response groups shown in Table 3-7.

The response groups were the same as those in Norman & Sigman (2009) who considered cases such as postal non-deliverables, vacant households, and other returned mail as ineligible. On the other hand, respondents were either those households that returned the mailed screener with a working phone number and completed the screener interview when contacted for the telephone interview, or that completed the screener interview when contacted through a matched telephone number.

Table 3-7. ABS sample screener response groups

Screener response status group	Description	Groups
A_E	Eligible	Telephone number available and household completed the screener interview
A_ENR	Eligible nonrespondent	Telephone number available but household refused screener interview Telephone number available but received after cut-off date Telephone number not available, household did not return mailed questionnaire Telephone number not available, household returned blank questionnaire Telephone number not available, household returned questionnaire without a telephone number
A_IN	Ineligible	Telephone number available but household reached does not match mailing address Telephone number available and household reported being outside Sonoma County Telephone number not available and return coded as postal non-delivery (PND) with new address, insufficient address, not deliverable as addressed, or vacant.

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

The ABS screener nonresponse adjusted household weight, $HHA8W_i$, was computed as

$$HHA8W_i = HHA8F_c * HHA7W_i ,$$

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

$$HHA8F_c = \begin{cases} \dfrac{\sum\limits_{i \in L\_E, L\_IN, L\_UNK} HHA7W_i \delta(c)}{\sum\limits_{i \in L\_E, L\_IN} HHA7W_i \delta(c)} & \text{If } i \in A\_ER \\ 0 & \text{If } i \in A\_ENR, A\_IN \end{cases} ,$$

where the groups *A_ER, A_NR*, and *A_IN* are defined in Table 3-7, and $\delta_i(c)$ is 1 if the number is in eligibility nonresponse adjustment cell c and is zero otherwise. The nonresponse adjustment cells were created by grouping geographically adjacent ZIP codes in the sample. Table B-1 in Appendix B (rows 9.1 through 9.4) shows the sums of weights before and after the screener nonresponse adjustment.

3.8. Multiple Telephone and Duplicate Respondent Adjustments

The screener interview for the landline sample collected information about the existence of additional landline telephone numbers and their use in the household. If more than one landline telephone number was used for residential purposes (not solely for business, fax or computer use, etc.), the household had 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 was adjusted to reflect the increased probability of selection. The multiple telephone-adjusted household weight, $HHA9W_i$, was 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 assumed that there was at most one additional residential-use landline telephone number in the household. Table B-1 in Appendix B (rows 10.1 through 10.5) identifies the sums of weights before and after the multiple telephone adjustment. This adjustment was not applied to the cell and ABS samples, where this adjustment factor was set to 1.

After adjusting the weights for the increased probability of selection due to multiple landline 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 number. 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 added to that of the completed screener. The duplicate respondent adjustment factor $ODF1_i$ was computed as:

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

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 stratum. In this case, the second duplicate respondent adjustment factor $ODF2_i$ was computed as:

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

The household weight adjusted for duplicate respondents, $HHA10W_i$, where the overall duplicate respondent factor $HHA10F_i$ adjustment factor was computed as $HHA10F0_i = ODF1_i * ODF2_i$ is

$$HHA10W_i = HHA10F0_i * HHA9W_i$$

The multiple telephone and duplicate respondent adjustments is not applicable to the ABS sample where the factor $HHA10F0_i = 1$ was set to 1 for all these cases. Table B-1 in Appendix B (rows 11.1 through 11.2) identifies the sums of weights before and after this adjustment.

4. ADULT WEIGHTING

A final weight was created for each adult who completed the adult extended interview.[7] The initial adult weight for the landline and surname samples was the product of the final household weight and the number of (eligible) adults in the household, that is, the reciprocal of the probability of selecting the adult from all adults in the household. For the cell phone sample, the initial weight was the product of the final household weight and the number of adults in the household if the cell phone was shared; if the cell phone was not shared, the initial adult weight equaled the final household weight. In subsequent steps, the initial adult weight was adjusted for nonresponse. Before raking the adult weights to known population control totals, the achieved landline and cell phone samples were poststratified to controls by telephone use. After this step, a composite weight combining the landline and cell phone samples 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. After raking, the ABS and telephone samples for Sonoma County were combined using a composite factor. In the last step, the weights of the combined sample were raked a second time to the same population control totals.

4.1. Adult Initial Weight

As described in *CHIS 2013-2014 Methodology Series: Report 1 - Sample Design*, one adult was sampled with equal probability from all adults in the household using the Rizzo method (Rizzo, Brick, & Park, 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

$$ADA0W_i = ADCNT_i \cdot HHA10W_i,$$

where $ADCNT_i$ is the total number of adults in household *i* for the landline and surname samples and for cell phone sample if there are adults that share the sampled phone, and $HHA10W_i$ is the multiple-telephone-adjusted weight described in the previous chapter.

This approach was also used for the ABS sample. On the other hand, this scheme for the cell phone sample assumes that, in cell phone households with more than one adult, each adult has a cell

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

phone (or shares a different cell phone) if the sampled cell phone is not shared. If the cell phone is shared, we assumed that all adults in the household share that 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, cell, and ABS samples.

4.2. Adult Nonresponse Adjustment

Regardless of the sample, 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 therefore 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, we classified extended interviews 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

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

The adult nonresponse adjusted weight, $ADA1W_i$, was computed as

$$ADA1W_i = ADA1F_c \cdot ADA0W_i ,$$

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

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

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 25 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 sums of weights before and after the nonresponse adjustment, for all 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
Language of the screener interview	1.	English
	2.	Other
Adult screener Respondent	1.	Sampled adult was screener respondent
	2.	Sampled adult was not screener respondent

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

4.3. Composite Weight

The next step in creating person weights was to combine the landline/list and cell samples. This process was the same for the adult, child, and adolescent weights. Before creating the composite weights, both samples were poststratified separately to control totals defined by type of telephone service (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 2014 for the U.S. West region. The poststratified person weight, $PPERW_j$ is computed as

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

where $PERW_j$ is the person weight (e.g., $ADA1W_i$ for adults) .and $TEL\_SERVICE\_CT_i$ is the control total by telephone service (landline only, cell phone only, or both services). Appendix B, Table B-5 (rows 1.1 through 1.3), shows the sums of weights before and after this adjustment.

Once the samples were poststratified, a composite weight that combined the landline and cell phone samples was created. Based on research in Brick, Flores Cervantes, Lee, & Norman (2011), the composite factor $\lambda = 0.50$ 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

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

where $PPERW_j$ is the poststratified person weight above. Table B-5 in Appendix B (row 2.1) shows the sums of weights before and after this adjustment. This adjustment was not applicable to the ABS sample, which was combined with the landline/cell phone sample at a later step (See Section 4.6).

4.4. Adult Trimming Factors

Before benchmarking the adult weights to the known total of adults in California in 2014, we examined the distribution of the composite weights to determine if there were weights with a large effect on either the estimates or their variances. 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, Ferraro, Wilson, & Brick, (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

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

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

The three statistics for the largest 20 weights in each self-reported 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 was based on the distribution of weights within sampling stratum.

The trimmed weight $TRMW_i$ is computed as

$$TRMW_i = TFACT_i * COMBW_i ,$$

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 in the combined landline and cell phone samples, 47 records were trimmed.[8] The trimming factor ranged between 0.2135 and 0.7893. In the adult extended interview in the ABS sample, two records were trimmed and the average trimming factor was 0.6242. Table B-5 (rows 2.1 and 3-1 to 3-3) in Appendix B shows trimmed weights by self-reported stratum and the sums of weights before and after trimming.

4.5. Adult Raked Weight

The next step in adult weighting of the combined landline/list/cell phone sample was raking the trimmed weights to population control totals to produce estimates consistent with the 2014 California Department of Finance (DOF) Population Estimates. Included in the raking adjustment is an undercoverage adjustment for adults in households without a telephone. The specific control totals and the method used to create them are described in Chapter 7. In the case of the ABS sample, the adult weights were raked to the 2014 Department of Finance (DOF) Population Estimates for Sonoma County.

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 (i.e., 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 belongs. 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,

[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 a revision 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.

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

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 combined landline and cell sample weights were raked to known control totals for California, while the ABS weights were raked to the control totals for Sonoma County. Table B-5 (rows 3.3 and 4.1 to 4.5) in Appendix B shows the sums of weights before and after this raking adjustment.

4.6. Combining the Adult Landline and ABS Samples

In the next step, the landline/list/cell sample was combined with the ABS sample using a composite factor applied to the weights as described in Section 2.5. The value of the composite factor was set to $\phi = 0.50$ because the combined landline/list/cell sample and ABS sample have similar sample sizes in Sonoma County. The expression of the composite weight, $COMB2W_j$, is

$$COMB2W_j = \begin{cases} RAKEDW_j & \text{If person } i \text{ lives outside Sonoma County} \\ \phi * RAKEDW_j & \text{If person } i \text{ was sampled in the landline/cell phone sample and lives in Sonoma County} \\ (1-\phi) * RAKEDW_j & \text{If person } i \text{ was sampled in the ABS sample and lives in Sonoma County} \end{cases} ,$$

In the last step, the combined landline/list/cell/ABS sample was re-raked to the same control totals for California (and Sonoma County) as in Section 4.5. The main reason was to remove the small differences in the sums of weights and the control totals result of the combination of the samples. The adult second raked weight, $RAKED2W_i$, can be expressed as

$$RAKED2W_i = CPMB2W_i \cdot \prod_{k=1}^{K} RAKED2F_{k_l} \ ,$$

where $RAKED2F_{k_l}$ is the second raking factor for dimension k, level l in which adult i belongs. As described before, the raking factors are derived so the following relationship holds for every raking dimension k, and level l,

$$CNT_{k_l} = \sum_i \delta(k_l)_i \cdot RAKED2W_i \ ,$$

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 were examined to determine the presence of large weights for trimming. However, no additional weights were trimmed. Table B-8 in Appendix B shows the sums of weights before and after the composite weight for combining the ABS and telephone samples and second raking adjustments.

5. CHILD WEIGHTING

A final child weight was created for all completed child extended interviews. Children (and adolescents) have been selected in both landline and cell samples since 2009. 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 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 use poststratification, composite weight, trimming, first raking, composite of landline/cell and ABS sample and second raking.

5.1. Household-Level Adjustment

The main difference between the child (and adolescent) weighting procedures and those for adults is that one adult was always sampled in the screener. Children and adolescents could be selected at the end of 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. This procedure was used for all samples except the cell phone sample. 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 children and adolescents selected during the child-first procedure 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 selected in screener?	Section G completed by adult?	Description
C1st	Yes	N/A	Households with child-first procedure
NC1stGC	No	Yes	Households without child-first procedure and section G was completed
NC1stGNC	No	No	Households without child-first procedure and section G was not completed

Source: UCLA Center for Health Policy Research, 2013-2014 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 $HHA10W_i$ was adjusted. We refer to this adjusted weight as the Section G adjusted household weight, $HHA11W_i$, and it is

$$HHA11W_i = HHA11F_c * HHA10W_i,$$

where

$$HHA11F_i = \begin{cases} \dfrac{\sum\limits_{i \in NC1stGC,\ NC1stGNC} HHA10W_i \delta(c)}{\sum\limits_{i \in NC1stGC} HHA10W_i \delta(c)} & \text{If } i \in NC1stGC \\ 0 & \text{If } i \in NC1stGNC \\ 1 & \text{If } i \in C1st \end{cases},$$

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 an additional adjustment to the household weight. This adjustment was also applied separately to the ABS sample. Table B-1 in Appendix B (rows 12.1 through 12.2) identifies the sums 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, 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 was the parent or legal guardian of the child). If the sampled adult did not have an associated child, then no

[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.

child was sampled even if there were children present in the household. See *CHIS 2013-2014 Methodology Series: Report 1 - Sample Design* for additional information on the within-household person selection process.

Since the selection of a child within a household depended 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

$$CHA0W_j = \frac{1}{CHPROB_j} HHA11W_i \,,$$

where $HHA11W_i$ 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 2013-2014 Methodology Series: Report 1 - Sample Design*, but we note here that within the same household children age 0 to 5 years had a measure of size twice that of children age 6 to 11 years. If the sampled adult had a spouse/partner living in the household and the spouse/partner of the sampled adult was 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) in Appendix B.

5.3. Other Child Weighting Adjustments

Adjustments to the child weights included one for extended interview nonresponse, telephone service poststratification, construction of composite weights, trimming influential weights, and raking to control totals. As in the adult weights, these adjustments were made separately for the landline/list, cell, and ABS samples. The child nonresponse adjustment is the same as the adult nonresponse adjustment described in Section 4.2, except for the definition of the adjustment cells. We created child nonresponse adjustment cells using two variables: 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 15 respondents, we collapsed cells to increase the number of respondents in each cell. Any cells still containing fewer than 15 respondents were collapsed combining age group. Table B-3 (rows 2.1 through 2.3) in Appendix B shows the number of sample records and sums 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 type. Table B-6 (rows 1.1 through 1.3) in Appendix B shows the sums 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 value of the composite factor (i.e., $\lambda = 0.50$) as in the adult sample. Table B-6 (row 2.1) in Appendix B shows the sums of weights after this adjustment.

The next step was to identify and trim large child weights in the landline/cell sample and ABS sample separately. The same 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 73 child weights, mainly in the landline/list/cell sample. The trimming factors range from 0.1294 to 0.8337. Appendix B Table B-6 (rows 2.1 and 3.1 through 3.3) shows the distribution of trimmed weights by sample type and the sum of the weights before and after applying the trimming factors.

In the next step, the landline/cell and ABS sample child weights were raked separately to population control totals to produce estimates consistent with the California Department of Finance 2014 population estimates. See Chapter 7 for the specific controls used. The expression for the raking adjustment was the same as that for adult weights described in Section 4.5. Then the cell phone/landline and ABS samples were combined using a composite factor as described in Section 2.5. The value of the composite factor was set to 0.5 and was only applicable to the respondents in Sonoma County. In the next two steps, the combined landline/list/cell phone/ABS sample was trimmed and then re-raked to the same control totals for California. The main reason was to remove the differences between the sums of weights and the control totals from combining the ABS and telephone samples. The expressions of the composite factor and second raking are similar to those described in Section 4.6 for the adult weights. As in the adult weights, no additional trimming of large weights was necessary after the second raking. Table B-9 in Appendix B shows the sums of weights before and after the composite weight for combining the ABS and telephone samples and second raking adjustments for the child weights.

6. ADOLESCENT WEIGHTING

In CHIS 2013-2014, adolescents were sampled and responded to the interview for themselves after parental permission was obtained 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, first raking, composite of landline/list/cell and ABS samples and second raking.

6.1. Initial Adolescent Weights

The steps for creating the adolescent weights are the same as those for creating the child weights described in Chapter 5. As in 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 computed as

$$TNA0W_j = \frac{1}{TNPROB_j} HHA11W_i \,,$$

where $HHA11W_i$ is defined in Chapter 5, and $TNPROB_i$ is computed in the same way as $CHPROB_i$ in Section 5.2. However, the measure of size is unity for all adolescents regardless of their age. Appendix B 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 in the same way as 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 sex of the adolescent and age group (12-14 and 15-17 years old) within sampling stratum. We inspected response rates separately by the two variables at the state level to determine the most important variables and the order of collapsing. After reviewing these rates, we created cells using sampling stratum, sex and age group. Cells containing fewer than 10 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/list and cell samples. As in the adult weights, child and adolescent weights from the landline/surname and cell phone samples were poststratified separately to control totals defined by telephone service. Appendix B, Table B-7 (rows 1.1 through 1.3) shows the sums of weights before and after this adjustment.

After poststratification, the adolescent landline and cell phone samples were combined using a composite factor. The value of the composite factor was for the adolescent sample was $\lambda = 0.50$. Appendix B, Table B-7 (row 2.1) shows the sums of weights before and after this adjustment. After the creating the composite weight, 76 influential weights were identified and trimmed with factors ranging from 0.0957 to 0.7853. Appendix B, Table B-7 (rows 2.1 and 3.1 through 3.3) shows 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 next step, the adolescent weights were raked to California DOF 2014 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 (see Section 4.5). Appendix B, Table B-7 (rows 3.3 and 4.2) show the sums of weights before and after raking.

In the next step, the landline/cell and ABS sample adolescent weights were then raked separately to population control totals separately to produce estimates consistent with the California Department of Finance 2014 population estimates. See Chapter 7 for the specific controls used. The expression for the raking adjustment was the same as that for adult weights described in Section 4.5. Then the cell phone/landline and ABS samples were combined using a composite factor as described in Section 2.5. The value of the composite factor was set to 0.5 and is only applicable to the respondents in Sonoma County. In the next two steps, the combined landline/cell phone/ABS adolescent were trimmed and then re-raked to the same control totals for California. The main reason was to remove the differences between the sums of weights and the control totals result of combining the ABS and telephone samples. The expression of the composite factor and second raking are similar to those described in Section 4.6 for the adult weights. As in the adult weights, no additional trimming of large weights was necessary after the second raking. Table B-10 in Appendix B shows the sums of weights before and after the composite weight for combining the ABS and telephone samples and second raking adjustments for the adolescent weights.

7. RAKING AND CONTROL TOTALS

This chapter describes the raking procedure and the development of control totals for CHIS 2013-2014. 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 12 dimensions used to rake the weights. The remaining sections describe the sources for deriving the control totals and how these control totals were derived.

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 of 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 & Rao (1979); Deville & Särndal (1992); and Kalton & 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 12 dimensions used in CHIS 2013-2014, 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 12 dimensions used in CHIS 2013-2014 are shown in Table 7-1. The first 8 dimensions and the 12th dimension in Table 7-1 were created by combining demographic variables (age, sex, race, and ethnicity) and different geographic areas (county, region or group of counties, region, 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. Section 7.3 describes this adjustment and the variables used to create the levels for this dimension. The raking dimensions for CHIS 2013-2014 are the same as those in CHIS 2011-2012 except for dimensions 4 and 7. Dimension 4 includes separate totals for the rural counties with larger sample sizes (Siskiyou, Tuolumne, and Calaveras Counties) and Dimension 7 includes separate counts for Non-Latino Japanese under 18 years old and 18 years old or older, who also have larger sample sizes in this CHIS cycle.

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

Dimension	Level	Description	Categories	
1	Region (R) (collapsed where necessary)	Age groups (3) x Sex (2)	11R	Under 12 years, male
			12R	Under 12 years, female
			21R	12 to 17 years, male
			22R	12 to 17 years, female
			31R	18 years or older, male
			32R	18 years or older, female
2	Region (R) (collapsed where necessary)	Age groups (9)	R1	Under 6 years
			R2	6 to 11 years
			R3	12 to 17 years
			R4	18 to 24 years
			R5	25 to 29 years
			R6	30 to 39 years
			R7	40 to 49 years
			R8	50 to 64 years
			R9	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	
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., Rural counties and Remainder of CA	SPAs (8), HRs (6), Remainder of CA (1)	0	Remainder of CA
			11	SPA 1 – Antelope Valley
			12	SPA 2 – San Fernando
			13	SPA 3 – San Gabriel
			14	SPA 4 – Metro
			15	SPA 5 – West
			16	SPA 6 – South
			17	SPA 7 – East
			18	SPA 8 – South Bay
			21	HR 1 – North Coastal
			22	HR 2 – North Central
			23	HR 3 – Central
			24	HR 4 – South
			25	HR 5 – East
			26	HR 6 – North Inland
			432	Siskiyou County
			441	Tuolumne County
			442	Calaveras County

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

Dimension	Level	Description		Categories
5	Region (R) (collapsed where necessary)	Race/ethnicity (7)	1	Under 12 years old (whole state)
			2	12 to 17 years old (whole state)
			1R	Latino 18 years old or older
			2R	Non-Latino White 18 years old or older
			3R	Non-Latino African American 18 years old or older
			4R	Non-Latino American Indian 18 years old or older
			5R	Non-Latino Asian 18 years old or older
			6R	Non-Latino Native Hawaiian 18 years old or older
			7R	Non-Latino Two or more races 18 years old or older
6	State	Race/ethnicity (7) x Age groups (2) x Gender (2) (collapsed where necessary)	111	Latino, Male, under 18 years
			112	Latino, Male, 18 years or older
			121	Latino, Female, under 18 years
			122	Latino, Female, 18 years or older
			211	Non-Latino White, Male, under 18 years
			212	Non-Latino White, Male, 18 years or older
			221	Non-Latino White, Female, under 18 years
			222	Non-Latino White, Female, 18 years or older
			311	Non-Latino African American, Male, under 18 years
			312	Non-Latino African American, Male, 18 years or older
			321	Non-Latino African American, Female, under 18 years
			322	Non-Latino African American, Female, 18 years or older
			411	Non-Latino American Indian, Male, under 18 years
			412	Non-Latino American Indian, Male, 18 years or older
			421	Non-Latino American Indian, Female, under 18 years
			422	Non-Latino American Indian, Female, 18 years or older
			511	Non-Latino Asian, Male, under 18 years
			512	Non-Latino Asian, Male, 18 years or older
			521	Non-Latino Asian, Female, under 18 years
			522	Non-Latino Asian, Female, 18 years or older
			611	Non-Latino Native Hawaiian, Male, under 18 years
			612	Non-Latino Native Hawaiian, Male, 18 years or older
			621	Non-Latino Native Hawaiian, Female, under 18 years
			622	Non-Latino Native Hawaiian, Female, 18 years or older
			711	Non-Latino Two or more races, Male, under 18 years
			712	Non-Latino Two or more races, Male, 18 years or older

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

Dimension	Level	Description		Categories
6	State	Race/ethnicity (7) x Age groups (2) x Gender (2) (collapsed where necessary)	721	Non-Latino Two or more races, Female, under 18 years
			722	Non-Latino Two or more races, Female, 18 years or older
7	State	Asian groups (5) x Age groups (2)	11	Non-Latino Chinese only, under 18 years
			12	Non-Latino Chinese only, 18 years or older
			21	Non-Latino Korean only, under 18 years
			22	Non-Latino Korean only, 18 years or older
			31	Non-Latino Filipino only, under 18 years
			32	Non-Latino Filipino only, 18 years or older
			41	Non-Latino Vietnamese only, under 18 years
			42	Non-Latino Vietnamese only, 18 years or older
			51	Other or non-Asian only, under 18 years
			52	Other or non-Asian only, 18 years or older
			61	Non-Latino Japanese only, under 18 years
			62	Non-Latino Japanese only, 18 years or older
8	Stratum (S) (collapsed where necessary)	Race/ethnicity (3) x Age groups (2)	S11	Latino, under 18 years
			S12	Latino, 18 years or older
			S21	Non-Latino White, under 18 years
			S22	Non-Latino White, 18 years or older
			S31	Non-Latino Non-White, under 18 years
			S32	Non-Latino Non-White, 18 years or older
9	Region (R) (collapsed where necessary)	Education (4)	R1	Not applicable (age < 18 years)
			R2	Less than High School
			R3	High School grad or GED recipient
			R4	At least some college
10	Region (R) (collapsed where necessary)	Person type (3) x # Adults in HH (3)	11R	Adult, 0 or 1 adult
			12R	Adult, 2 adults
			13R	Adult, 3 or more adults
			21R	Child, 0 or 1 adult
			22R	Child, 2 adults
			23R	Child, 3 or more adults
			31R	Teen, 0 or 1 adult
			32R	Teen, 2 adults
			33R	Teen, 3 or more adults
11	Region (collapsed where necessary)	Non-telephone dimension		See Table 7-3. Dimension 11, non-telephone adjustment cell definition for CHIS 2013-2014
12	Region (7) x Stratum (S)	Person type (3)	RSS1	Child
			RSS2	Teen
			RSS3	Adult

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

Before raking the weights, dimensions with levels or cells with fewer than 50 respondents were collapsed with “adjacent” levels to form larger cells. Cells in dimensions defined at the stratum level were collapsed within the geographic regions shown in Table 7-2. Cells of dimensions defined at the region level were collapsed across regions if the raking cells 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 the smaller 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.

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, 2013-2014 California Health Interview Survey.

7.3. Non-Telephone Raking Dimension

The main components of CHIS 2013-2014 included both landline and cell phone samples so respondents with at least one telephone type have a chance of being selected. However, there is potential for bias from undercoverage from households without any telephone service if there are differences in characteristics of the persons residing in households with telephones and those without. CHIS 2013-2014 includes a nontelephone adjustment focused on reducing the potential bias introduced by exclusion of nontelephone households from the survey. This adjustment was carried out through a raking dimension at the person level (dimension 11). The control totals were derived for the same cells using the 2014 California Department of Finance (DOF) Population Estimates and Population Projections and the 2011-

2013 American Community Survey public use micro data file (ACS-PUMS) (U.S. Census Bureau, 2014). Table 7-3 shows the definition of the cells of dimension 11.

Table 7-3. Dimension 11, non-telephone adjustment cell definition for CHIS 2013-2014

Dimension 11 levels	Stratum	Household tenure	Age in years	Educational attainment	Number of adults in the household
1R101	Region (R)	Own	0 to 17	NA	0 or 1
2R101		Rent	0 to 17	NA	0 or 1
1R102		Own	0 to 17	NA	2 or more
2R102		Rent	0 to 17	NA	2 or more
1R210		Own	18 to 30	Up to high school	NA
1R310		Own	31 to 64	Up to high school	NA
1R410		Own	65 and older	Up to high school	NA
1R220		Own	18 to 30	Greater than high school	NA
1R320		Own	31 to 64	Greater than high school	NA
1R420		Own	65 and older	Greater than high school	NA
2R210		Rent	18 to 34	Up to high school	NA
2R311		Rent	35 and older	Up to high school	0 or 1
2R312		Rent	35 and older	Up to high school	2 or more
2R220		Rent	18 to 34	Greater than high school	NA
2R321		Rent	35 and older	Greater than high school	0 or 1
2R322		Rent	35 and older	Greater than high school	2 or more

Source: UCLA Center for Health Policy Research, 2013-2014 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 combined landline/list and cell phone samples. The overall adjustment factors were computed as the ratio of the control total to the sums of weights before raking. The factors in the table can only be compared to the CHIS 2009 and CHIS 2011-2012 weights because these weights have similar weighting adjustments (Sections 4.3, 5.3, and 6.2) which were not used in earlier cycles. Further, because of the telephone use poststratification, 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	1.040	0.975	1.018
Female	0.964	0.984	1.073
Age group			
Under 5 years		1.082	
6-11 years		0.896	
12-17 years			1.044
18-24 years	1.024		
25-29 years	1.335		
30-39 years	1.199		
40-49 years	1.032		
50-64 years	0.848		
65 years and over	0.905		
Race/Ethnicity [a]			
Latino	1.054	1.077	1.161
Non-Latino			
White alone	0.934	0.752	0.853
African American alone	0.961	1.011	0.985
American Indian/Alaska Native alone	0.711	0.859	1.416
Asian alone	1.142	1.335	1.578
Native Hawaiian and Other Pacific Islander alone	1.411	0.621	0.597
Two or more races	0.904	0.935	0.961
Non-Latino Asian ethnic groups			
Chinese only	1.072	1.073	1.228
Korean only	0.743	1.684	1.886
Filipino only	1.778	2.367	2.698
Vietnamese only	0.861	1.376	1.431
Japanese only	1.537	3.666	3.688
Educational Attainment			
Not applicable (age < 18 years)		0.979	1.044
Less than High School,	0.970		
High School grad or GED recipient,	1.047		
Some college	0.996		
College degree or above	0.986		
Household Tenure [a]			
Owner	1.149	1.046	1.149
Renter	0.851	0.933	0.934
Number of adults in the household [b]			
One	0.825	1.173	1.020
Two	0.951	0.915	1.080
Three or more	1.140	1.080	1.005
Number of children in the household [b]			
None	0.973		1.016
One	1.073	1.071	1.084
Two or more	1.130	0.935	1.082

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

Characteristic	Adult	Child	Adolescent
Number of adolescents in the household [b]			
None	0.996	0.989	
One	1.027	0.945	1.118
Two or more	1.023	0.966	0.955

[a] OMB race ethnicity

[b] Person level estimate by type of household

Source: UCLA Center for Health Policy Research, 2013-2014 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 telephone surveys. The factors also suggest that younger adults (under 50 years old), and adults who own their home, adults in households with three or more adults, adults in households with at least one child or adolescent are harder to reach. The factors for the Latino, non-Latino Asian and non-Latino Native Hawaiian and Other Pacific Islander groups are also all larger suggesting potential undercoverage.

7.5. Sources Used to Produce the Control Totals for CHIS 2013-2014

Since the beginning of CHIS, considerable thought was given to the choice of data for the primary source of the control totals. It is desirable to use an up-to-date source available at the county or sub-county level with separate totals by age, race, sex, and ethnic groups. In 2003 the California Department of Finance (DOF) Population Projections were selected as the primary source for control totals for CHIS supplemented by other sources. The DOF estimates have been the main source of control totals since CHIS 2003. These files are described in the following sections.

7.5.1. California Department of Finance Population Predictions and Estimates

Based on discussions with UCLA, the 2014 California DOF Population Projections poststratified to 2014 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 2013-2014 (State of California, Department of Finance, 2013). The population projections are available at the county level by race, ethnicity, gender and single age for each year and are projected 50 years into the future. The projections are revised after each decennial census.

The 2014 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. Definition of counts available in the 2014 California DOF population projections files* The DOF population projections used the 2010 Census counts as the baseline. The DOF uses a baseline cohort-component method to project population estimates based on fertility/mortality rates and life expectancy by different race-ethnic groups and age cohorts. Special populations (those in prisons, colleges, and military installations) that have very different demographic and behavioral characteristics from the household population were removed from the baseline and projected separately. However, the DOF files held most of the special populations only for 2010. This factor played an important role in the assumptions made when removing the population living in group quarters from the control totals in CHIS 2013-2014 as described in Section 7.6.1.

Table 7-5. Definition of counts available in the 2014 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 White alone Latino African American alone Latino American Indian/Alaska Native alone Latino Asian alone Latino Native Hawaiian and Other Pacific Islander alone Latino Two or more races 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.

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 a sixth category, “some other race,” consistent with the 2010 Census data collection method. Recoding of “other race” for CHIS 2013-2014 largely followed Census procedures (see *CHIS 2013-2014 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. In order to reduce the number of imputations of “other race” respondents, a variable combining ethnicity with OMB race was proposed and approved by UCLA. The creation and imputation of this variable, OMBSRREO, is described in Section 8.4.2

The DOF also provides Population Estimates (State of California, Department of Finance, 2014) for current and previous years. 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 before the projected year and the latter after at the end of the estimated year. Therefore, the distributions of the DOF Population Estimates are more representative of the population. The disadvantage is that the population estimates are only available for the total population at the county level.

Both the DOF population projections and estimates include the population living in group quarters. Since the target population in CHIS 2013-2014 excludes persons living in group quarters, these persons were removed from the DOF population projections. The Census 2010 files were used to estimate the proportion of persons in group quarters, and these proportions were removed from the DOF estimates.

7.5.2. Census 2010 Files

As in previous cycles of CHIS, the DOF population totals were adjusted to remove the population living in group quarters who was not eligible for the survey. The 2010 Census Summary File 1 of SF1 (U.S. Census Bureau, 2012a) was used to compute the proportion of persons living in group quarters. Section 7.6.1 describes the details of this process

The 2010 SF1 was 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 2010 Census files. This assumes that the proportion in these areas with respect to the county did not change between 2010 and 2014.[11] The Los Angeles SPAs and San Diego Health Regions were both defined in terms of Census Tracts.

The 2010 Census Summary File 2 or SF2 (U.S. Census Bureau, 2012c) was used to compute the control totals for the Asian ethnic groups in dimension 7. The 2010 Census Modified Race File (U.S. Census Bureau, 2012b) was used to adjust the Census SF1 files to produce totals that include “other race” as a separate race category not found in the DOF files.

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 2011-2013 California ACS public use micro data file (PUMS) (U.S. Census Bureau, 2014) 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 2014 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 after assigning each Public Use Microdata Areas (PUMAs) to a region in California. Applying the 2011-2013 factors assumed that there were no changes in the population proportions between 2011-2013 and 2014 for these variables.

7.5.4. The National Health Interview Survey

The National Health Interview Survey (NHIS) is one of the major data collection programs of the National Center for Health Statistics (NCHS), part of the Centers for Disease Control and Prevention (CDC), and has been conducted since 1957. The NHIS is an in-person survey where sampling and interviewing are conducted continuously throughout the year. The survey collects information about household telephone service 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 for January to June 2014 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

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

NHIS does not produce current estimates at the state level, we use the estimates for the West region. Table 7-6 shows the percentages that were applied to the DOF totals to derive the control totals for poststratification for telephone use (see Section 4.3).

Table 7-6. NHIS proportions of telephone use by person type

Person type	Telephone service	Proportion
Adults	Landline only	0.073
	Cell phone and landline	0.450
	Cell phone only	0.477
Children and adolescents	Landline only	0.040
	Cell phone and landline	0.442
	Cell phone only	0.518

Source: 2014 Q1 and Q2 NHIS special tabulation.

7.6. Producing the Control Totals for CHIS 2013-2014

As with previous cycles of CHIS, the derivation of the control totals was a challenging task in 2014. It involved the selection of the sources of control totals, determining the number of and levels of dimensions, and computing the control totals. It also had an impact on the set of variables to be imputed. In CHIS 2013-2014, there were 12 raking dimensions. Deriving the control totals for each dimension independently could lead to inconsistencies between totals across the dimensions, which could cause problems in raking.

To overcome these difficulties, we used the procedure developed in 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 2010 and 2014, the Census 2010 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 only available by stratum (44) × age group 1 (3) × sex (2), stratum (44) × age group 2 (2) × sex (2) × race (7), and stratum (44) × age group 2 (2) × sex (2) × ethnicity (3) as defined in Table 7-6. The Census 2010 files did not include as many detailed group quarters as in the Census 2000 files. For example, the population in group quarters by single age was not available. As a result, the process to remove the population in group quarters was modified based on the limited totals. In the modified procedure, it was assumed that the distribution of the population in group quarters is uniform among three age groups (less than 18 years old, 18 to 64 years old, and 65 years old or older). For example, if 1.56 percent of persons 65 or older are in group quarters, 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 7-7 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.

Table 7-7. Definition of levels of variables for group quarters populations in the Census 2010 SF1 file

Characteristics	Available counts
Stratum (44)	Counties or combinations of multiple counties defined in CHIS 2013-2014
Age group1 (3)	Less than 18 years old 18 to 64 years old 65 years old or older
Age group2 (2)	Less than 18 years old 18 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 2010.

Under these assumptions, we computed the percentage of the population not living in group quarters in 2010. A file with 2010 population totals, T_{rc}^{2100}, was created by summarizing the 2010 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 7-8) 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 7-8. Age levels corresponding to the cross-tabulation of the DOF data files and the definition of the raking dimensions

Age group (18)	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

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

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 $T_{rc}^{2010\,\overline{GQ}}$ as the 2010 population total that excludes the population in group quarters in cell *rc*. The totals $T_{rc}^{2010\,\overline{GQ}}$ were computed by raking the totals T_{rc}^{2010} to three control totals for the population not living in group quarters. Let $D1_m^{2010\,\overline{GQ}}$ be the control total for the first raking dimension computed as

$$D1_m^{2010\,\overline{GQ}} = D1_m^{2010} - D1_m^{2010\,GQ},$$

where $D1_m^{2010}$ is the 2010 total population, $D1_m^{2010\,GQ}$ is the 2010 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 $D2_n^{2010\,\overline{GQ}}$ be the control total for the second raking dimension for cell *n* defined as the cross-tabulation of strata(44) × ethnicity(3) × age group 1 (3) × sex(2) as in the SF1. Similarly, let $D3_p^{2010\,\overline{GQ}}$ be the control total for the third raking dimension for cell *p*, where *p* is defined as the cross-tabulation of strata (44) × age group 2 (2) as in the SF1.

Note that $D1_m^{2010\,GQ}$, $D2_n^{2010\,GQ}$, and $D3_p^{2010\,GQ}$ are the 2010 population totals living in group quarters available in the SF1 file. By using raking we ensured that all totals, $T_{rc}^{2010\,\overline{GQ}}$, were consistent and they summed to the control totals.

After raking, the proportion of the 2010 population not living in group quarters in cell *rc* was computed as

$$p_{rc}^{2010\,\overline{GQ}} = \frac{T_{rc}^{2010\,\overline{GQ}}}{T_{rc}^{2010}}.$$

Assuming that the proportion of the population not living in group quarters did not change between 2010 and 2014 within cell *rc*, the proportion $p_{rc}^{2010\,\overline{GQ}}$ could be used to compute $T_{rc}^{2014\,\overline{GQ}}$ defined as the 2014 total population not living in group in cell *rc*, as

$$T_{rc}^{2014\,\overline{GQ}} = p_{rc}^{2010\,\overline{GQ}} * T_{rc}^{2014},$$

where T_{rc}^{2014} is the 2014 total population from the 2014 California DOF file in cell *rc*. However, T_{rc}^{2014} could not be computed using the DOF file due to differences in race categorization between the SF1 and the DOF projection. Instead, the 2014 population estimates, $T_{sc}^{2014\,OMB}$, 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 7-9, and *c* is defined as before.

Table 7-9. OMB race categories available in the California DOF files

$race_{OMB}(s)$	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

Source: State of California, Department of Finance.

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),

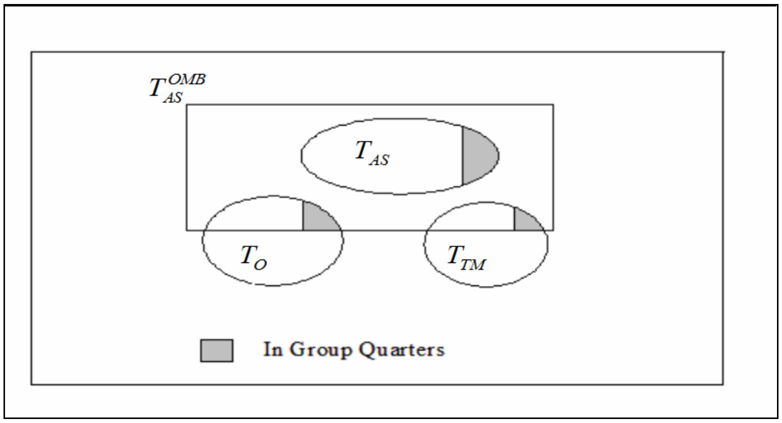
$$T_{ASc}^{OMB} = T_{ASc} + p_{AS\_O}^{OMB} * T_{Oc} + p_{AS\_TM}^{OMB} * T_{TMc} ,$$

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 illustrates 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,

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

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 2010 and 2014 the proportion was computed as

$$p_{sc}^{2014\,\overline{GQ}} = p_{sc}^{2010\,\overline{GQ}} = \frac{T_{rc}^{2010\,\overline{GQ}}}{T_{rc}^{2010}}.$$

The proportion $p_{sc}^{2014\,\overline{GQ}}$ was used to compute the 2014 total population not living in group quarters in cell sc, $T_{sc}^{2014\,OMB\,\overline{GQ}}$, defined using the OMB race categories, as follows:

$$T_{sc}^{2014\,OMB\,\overline{GQ}} = p_{sc}^{2014\,\overline{GQ}} * T_{sc}^{2014\,OMB} = \frac{T_{rc}^{2010\,\overline{GQ}} * T_{sc}^{2014\,OMB}}{T_{rc}^{2010}}.$$

where T_{rc}^{2010} is computed using the SF1 file, $T_{sc}^{2014\,OMB}$ using the 2014 DOF file and $T_{rc}^{2010\,\overline{GQ}}$ is the 2010 population in cell *rc* not in group quarters, as defined earlier. The 2012 total population not living in group quarters in California is computed as

$$T^{2014\,\overline{GQ}} = \sum_s \sum_c T_{sc}^{2014\,OMB\,\overline{GQ}}.$$

Table 7-10 shows the total population in the 2014 DOF file and the estimated total (and percentage) of the population living in group quarters.

Table 7-10. Estimated population in California in 2014 by group quarter status

Type	Population	%
In group quarters	916,946	2.38
Not in group quarters	37,582,432	97.62
Total	38,499,378	100.00

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

7.6.2. Computing the Control Totals

The totals $T_{sc}^{2014\,OMB\,\overline{GQ}}$ 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 demographic totals but for using the Asian only total and the Census 2010 SF2 file. The percentages of the Asian groups by ethnicity (Latino, non-Latino) were computed using the 2010 SF2 file. It was assumed that there were no changes in the distribution of the Asian groups between 2010 and 2014. These percentages were applied to the 2014 DOF projections.

The creation of dimension 4, defined by SPAs in Los Angeles County, Health and Human Services Agency (HHSA) Service Regions in San Diego County, and rural counties (Siskiyou, Tuolumne, and Calaveras Counties), used information from the Census 2010 SF1. The Los Angeles County Department of Health (LACDH) produced a listing of Census tracts by SPA. The 2010 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 2012 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 2011-2013 ACS-PUMS and then applied to the 2014 DOF population total (excluding group quarters). The underlying assumption was that there were no changes in the distribution of the population between the 2011-2013 ACS and 2014.

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 2013-2014. *CHIS 2013-2014 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 2013-2014 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 2013-2014. We 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 2013-2014. The same set of variables was imputed in CHIS 2011-2012 except for the variable OMBSRASO which includes an additional level for Japanese not imputed in previous cycles of CHIS. 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 estimates due to 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, KAM33	Telephone service
HASLANDLINE	Landline telephone service in household	AN6, AN7	Telephone service

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

In CHIS 2013-2014 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 2013-2014, 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 2013-2014 is very small across all samples (landline, surname list, cell phone, and ABS) 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 two adults, two children, and one adolescent in the landline sample were missing. These five cases were imputed randomly.

Age was imputed in 98 cases in CHIS 2013-2014 across all samples (91 adults and seven children). 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/Lists					
Adult	32,007	2	0.01	82	0.26
Child	4,214	2	0.05	6	0.14
Adolescent	1,756	1	0.06	0	0.00
Total	37,977	5	0.01	88	0.23
Cell phone					
Adult	7,752	0	0.00	8	0.10
Child	1,256	0	0.00	1	0.08
Adolescent	482	0	0.00	0	0.00
Total	9,490	0	0.00	9	0.09
ABS					
Adult	481	0	0.00	1	0.21
Child	42	0	0.00	0	0.00
Adolescent	15	0	0.00	0	0.00
Total	538	0	0.00	1	0.19
Overall Total	48,005	5	0.01	98	0.20

Source: UCLA Center for Health Policy Research, 2013-2014 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 323 missing responses (0.67 percent), and educational attainment had 260 missing responses (0.65 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, 2013-2014 California Health Interview Survey.

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

	Adult interviews							
	Sample type							
	Landline/lists		Cell phone		ABS		All samples	
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Self-reported Education Attainment								
Under 18 years of age	NA	NA	NA	NA	NA	NA	NA	NA
Less than HS, 18 years of age or older	41	1.24	11	1.26	0	0.00	52	1.22
High School (or equivalent), 18 years of age or older	51	0.77	9	0.49	0	0.00	60	0.70
Some college, 18 years of age or older	65	0.72	5	0.22	1	0.76	71	0.62
BS and above, 18 years of age or older	73	0.57	4	0.15	0	0.00	77	0.48
Total	230	0.72	29	0.38	1	0.21	260	0.65

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

	Adult interviews							
	Sample type						All Samples	
	Landline/Lists		Cell phone		ABS			
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Self-reported Household Tenure								
Owner	169	0.64	19	0.40	2	0.50	190	0.60
Renter	97	0.86	36	0.77	0	0.00	133	0.82
Total	266	0.71	55	0.58	2	0.37	323	0.67

Source: UCLA Center for Health Policy Research, 2013-2014 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 2014 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 2010 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 2010 include six race categories, the Census Bureau released a special file called Modified Race Data Summary file (MRDSF) with 2010 population counts by the five OMB race categories (U.S. Census Bureau, 2012b). 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 2010 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 2010, 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 4,612 persons (9.61 percent) in CHIS 2013-2014. 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 2013-2014 as in Census 2010, because the number of CHIS cases in the geographic area (i.e., stratum) by Latino origin[12] 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.

Table 8-5. OMB 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 86 persons (0.19 percent) who self-reported as non-Latino and "some other race" alone[13] in 2013-2014. 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 4,516 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 2013-2014. 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

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

[13] This includes records imputed as non-Latino "other" from the regular CHIS 2013-2014 race imputation.

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.

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/Lists	62	0.16	160	0.42
Adult	57	0.18	125	0.39
Child	1	0.02	20	0.47
Adolescent	4	0.23	15	0.85
Cell phone	17	0.18	23	0.24
Adult	15	0.19	16	0.21
Child	1	0.08	4	0.32
Adolescent	1	0.21	3	0.62
ABS	0	0.00	3	0.56
Adult	0	0.00	2	0.42
Child	0	0.00	1	2.38
Adolescent	0	0.00	0	0.00
Total	79	0.16	186	0.39

* At least one value of race was imputed.

Source: UCLA Center for Health Policy Research, 2013-2014 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 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 a 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	1,255	2.61	906	2.25	226	4.10	123	5.46
African American alone	34	0.07	21	0.05	7	0.13	6	0.27
Asian alone	37	0.08	27	0.07	5	0.09	5	0.22
American Indian/ Alaska Native alone	74	0.15	54	0.13	13	0.24	7	0.31
Pacific Islander alone	10	0.02	4	0.01	2	0.04	4	0.18
Other race alone	943	1.96	690	1.71	123	2.23	130	5.77
Two or more races	45	0.09	30	0.07	8	0.15	7	0.31
Total	2,398	5.00	1,732	4.30	384	6.97	282	12.52
Self-reported Ethnicity								
Latino	43	0.09	25	0.06	9	0.16	9	0.40
Non-Latino	143	0.30	118	0.29	16	0.29	9	0.40
Total	186	0.39	143	0.36	25	0.45	18	0.80
Completed interviews	48,005	100.00	40,240	100.00	5,512	100.00	2,253	100.00

Source: UCLA Center for Health Policy Research, 2013-2014 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 2010. Since the OMB assignment is done using the imputed regular single-race variables, all sampled persons have nonmissing race 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, 86 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	0	0.00	0	0.00	0	0.00	0	0.00
2. Non-Latino White alone	63	0.13	57	0.14	2	0.04	4	0.18
3. Non-Latino African American alone	5	0.01	5	0.01	0	0.00	0	0.00
4. Non-Latino Asian alone	2	0.00	2	0.00	0	0.00	0	0.00
5. Non-Latino American Indian/ Alaska Native alone	5	0.01	5	0.01	0	0.00	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	4	0.01	3	0.01	0	0.00	1	0.04
Total	86	0.18	72	0.18	2	0.04	5	0.22
Completed interviews	48,005	100.00	40,240	100.00	5,512	100.00	2,253	100.00

Source: UCLA Center for Health Policy Research, 2013-2014 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 2013-2014, 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. The additional level (i.e., OMBSRASO=6, Non-Latino Japanese alone) was added for 2013-2014 due to the increase sample size for this group.

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
6	Non-Latino Japanese alone

Source: UCLA Center for Health Policy Research, 2013-2014 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), six 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
SRJP	Self-reported Japanese
SRASO	Self-reported Other Asian ethnic group

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

The missing values of the OMB Asian group variables (SRCH, SRPH, SRKR, SRVT, SRJP, 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, SRJP, 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, SRJP, 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	13	0.03	13	0.03	0	0.00	0	0.00
2. Non-Latino Korean	4	0.01	3	0.01	1	0.02	0	0.00
3. Non-Latino Filipino	2	0.00	2	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	2	0.00	2	0.00	0	0.00	0	0.00

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

	Total		Extended interview type					
			Adult		Child		Adolescent	
OMB Asian group (OMBSRASO)	Imputed count	%	Imputed count	%	Imputed count	%	Imputed count	%
6. Non-Latino Japanese	51	0.11	32	0.08	11	0.20	8	0.36
Total	73	0.15	53	0.13	12	0.22	8	0.36
Completed interviews	48,005	100.00	40,240	100.00	5,512	100.00	2,253	100.00

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

8.4.4. Telephone Service

In CHIS 2013-2014, the weights of the landline, lists, and cell phone samples were poststratified to totals for telephone service (i.e., landline only, dual users, cell only). Information about the presence of a cell phone for respondents in the line sample was collected through questions AM33 and KAM33 (Respondent has a working cell phone). Similarly, information about the presence of a landline in the cellphone sample is collected through questions AN6 (landline phone in household) and AN7 (landline phone personal or business use). These items are used to create the variables HASCELL and HASLANDLINE which then are used to create the poststratification cells for the telephone use adjustment (see Section 4.3). These variables were imputed at the household level and all competed interviews within the household shared the same values. Hot-deck imputation was used to impute missing values for these two variables. Similarly to the imputation of household tenure, the search algorithm CHAID 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. The same variables used to impute for household tenure listed in Table 8-3 were used to impute the variables related to telephone service. Table 8-12 shows the distribution of the imputed cases by sample type.

Table 8-12. Counts and percentages of imputed telephone type*

	Total		Sample type			
			Landline/list		Cell phone	
	Count	Percentage	Count	Percentage	Count	Percentage
Has landline						
Yes	371	0.90	318	0.96	53	0.68
No	25	0.06	0	0.00	25	0.32
Total	396	0.97	318	0.96	78	1.01

Table 8-12. Counts and percentages of imputed telephone type* (continued)

	Total		Sample type			
			Landline/list		Cell phone	
	Count	Percentage	Count	Percentage	Count	Percentage
Has cell phone						
Yes	338	0.82	260	0.78	78	1.01
No	58	0.14	46	0.14	0	0.00
Total	396	0.97	306	0.92	78	1.01
Completed households	41,015	100.00	33,263	100.00	7,752	100.00

* Counts exclude ABS sample where these variables were not imputed.

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

8.4.5. Self-Reported County and Self-Reported Stratum

In CHIS 2013-2014, 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 8-13 shows the variables used in the geocoding process.

Table 8-13. Variables used in geocoding

Variable	Description	Source
AH42	County of residence (self-reported)	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-reported)	Adult questionnaire
AO2ADDR	Street address (self-reported)	Adult questionnaire
AO2CITY	City (self-reported)	Adult questionnaire
AM8	Street name of residence (self-reported)	Adult questionnaire
AM9	Street name of nearest cross street (self-reported)	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, 2013-2014 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 8-14 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 8-14. 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 8-14. Distribution of self-reported strata and sampling strata for the landline/list samples

	Stratum	Sampling stratum	Self-reported stratum	Net agreement ratio
1	Los Angeles	5,745	5,746	1.00
2	San Diego	3,419	3,419	1.00
3	Orange	1,783	1,750	0.98
4	Santa Clara	1,098	1,133	1.03
5	San Bernardino	1,066	1,068	1.00
6	Riverside	1,217	1,239	1.02
7	Alameda	958	924	0.96
8	Sacramento	909	915	1.01
9	Contra Costa	613	657	1.07
10	Fresno	501	499	1.00
11	San Francisco	612	605	0.99
12	Ventura	490	501	1.02
13	San Mateo	457	440	0.96
14	Kern	443	448	1.01
15	San Joaquin	382	382	1.00
16	Sonoma	867	872	1.01
17	Stanislaus	413	398	0.96
18	Santa Barbara	401	393	0.98
19	Solano	418	402	0.96
20	Tulare	403	401	1.00
21	Santa Cruz	403	388	0.96
22	Marin	407	413	1.01
23	San Luis Obispo	390	397	1.02

Table 8-14. Distribution of self-reported strata and sampling strata for the landline/list samples (continued)

	Stratum	Sampling stratum	Self-reported stratum	Net agreement ratio
24	Placer	347	410	1.18
25	Merced	383	398	1.04
26	Butte	387	397	1.03
27	Shasta	398	390	0.98
28	Yolo	396	400	1.01
29	El Dorado	433	389	0.90
30	Imperial	387	388	1.00
31	Napa	405	420	1.04
32	Kings	401	399	1.00
33	Madera	410	406	0.99
34	Monterey	384	396	1.03
35	Humboldt	394	390	0.99
36	Nevada	423	403	0.95
37	Mendocino	418	419	1.00
38	Sutter	399	407	1.02
39	Yuba	398	372	0.93
40	Lake	394	390	0.99
41	San Benito	403	403	1.00
42	Colusa, Glenn, Tehama	318	329	1.03
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Trinity	206	211	1.02
43.2	Siskiyou	435	421	0.97
44	Alpine, Amador, Inyo, Mariposa, Mono	124	140	1.13
44.1	Tuolumne	390	387	0.99
44.2	Calaveras	460	433	0.94

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

Table 8-15 shows the distribution of adult respondents by sampling stratum, self-reported stratum agreement ratios for the cell phone sample. The geographic area covered by a single landline stratum is not exactly the same as the area covered by equivalent cell phone stratum despite having the same name. As a result, the NARs are not comparable between the landline and cell phone sample. The table shows that the net agreement is more variable than the same ratio for the landline sample. The reason is the mobility of the cell phones; the area where they are sampled does not always match the area where the respondent resides.

Table 8-15. Distribution of self-reported area and sampling strata for the cell phone sample

Sampling stratum	Counties covered	Sampling stratum	Self-reported area	Net agreement ratio
1	Los Angeles	1,448	1,432	0.99
2	San Diego	846	832	0.98
3	Orange	420	426	1.01
4	Santa Clara	281	284	1.01
5	San Bernardino	245	260	1.06
6	Riverside	235	309	1.31
7	Alameda	257	238	0.93
8	Sacramento	216	238	1.10
9	Contra Costa	111	168	1.51
10	Fresno	120	139	1.16
11	San Francisco	195	162	0.83
12	Ventura	115	121	1.05
13	San Mateo	93	114	1.23
14	Kern	99	111	1.12
15	San Joaquin	95	97	1.02
16	Sonoma	90	115	1.28
17	Stanislaus	87	88	1.01
18	Santa Barbara	92	93	1.01
19	Solano	96	94	0.98
20	Tulare	97	103	1.06
21	Santa Cruz	109	101	0.93
22	Marin	103	87	0.84
23	San Luis Obispo	89	93	1.04
24	Placer	91	90	0.99
25	Merced	111	107	0.96
26	Butte	91	115	1.26
27	Shasta	108	99	0.92
28	Yolo	118	104	0.88
29	El Dorado	102	98	0.96
30	Imperial	135	101	0.75
31	Napa	125	106	0.85
32	Kings	113	99	0.88
33	Madera	120	102	0.85
34	Monterey	89	95	1.07
35	Humboldt	97	102	1.05
36	Nevada	88	89	1.01
37	Mendocino	134	106	0.79
38	Sutter	249	118	0.47
39	Yuba	13	91	7.00
40	Lake	94	96	1.02
41	San Benito	126	93	0.74
42	Colusa, Glenn, Tehama	71	83	1.17

Table 8-15. Distribution of self-reported area and sampling strata for the cell phone sample (continued)

Sampling stratum	Counties covered	Sampling stratum	Self-reported area	Net agreement ratio
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	68	79	1.16
44	Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne	70	74	1.06

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

9. VARIANCE ESTIMATION

This chapter describes the methods for and results of computing sampling errors for CHIS 2013-2014 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 2013-2014. 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 2013-2014 estimates using commercial and open source 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, & 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

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

where $\hat{p}$ is the estimated proportion, $v\left(\hat{p}\right)_{SRS}$ is the variance estimate of the estimated proportion assuming a simple random sample, and $v\left(\hat{p}\right)_{COMPLEX}$ 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 2013-2014, 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 2013-2014, because it would not have achieved the sample sizes for the domains of interest, in particular at the county/stratum level, for given resources. The design effects calculated from the CHIS 2013-2014 data indicate that the sample design used in the survey needs to be taken into account when analyzing the data.

In CHIS 2013-2014, 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 effect on the variability.

The tables in the following sections show the *DEFF*s and *DEFT*s for selected items from 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, & 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 also had disproportionate sampling because it was also allocated by county. However, when the samples are combined, persons in cell only households were subsampled.

- **Subsampling.** Inactive cell phones were sampled in the second year of data collection. This subsampling increased the design effect of the cell phone sample for statewide estimates and county estimates.

- **Within-Household Subsampling.** For all samples only one adult was selected in each household. 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 2013-2014 weights combine samples from overlapping domains that were sampled at different rates.

9.1.1. Design Effect for the Combined Sample Weights

Table 9-1 to Table 9-3 present the *DEFF*s and *DEFT*s of the adult, child and adolescent interviews, respectively, for the combined landline/list/cell sample. The first panel in each table 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 44 items selected from the adult interview, 31 items from the child interview, and 28 items from the adolescent interview. The variables include health characteristics such as general health rating, diagnosis (i.e., asthma, diabetes, high blood pressure, heart failure/congestive, heart disease, difficulty learning and remembering, child visited emergency room, felt nervous, had psychological or emotional counseling), lifestyle (smoking and alcohol, go to the park, had fast food), preventive medicine (mammogram, blood test, flu vaccine, delayed medical care, childcare), health insurance (insured, employer health insurance, other government health 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 1.63 and 1.83. This implies that for 75 percent of the strata the standard error of the estimates is about 63 to 83 percent greater than the expected standard error of a simple random sample. The average *DEFT* for the state estimates is 1.99. This is larger than the county-level *DEFT*s of 33 counties because most 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	4.10	4.24	6.14	0.16	1.99
Los Angeles	2.98	3.05	4.57	0.14	1.70
San Diego	3.49	3.47	5.80	1.02	1.85
Orange	3.48	3.41	6.91	1.32	1.84
Santa Clara	3.46	3.25	8.18	0.60	1.83

Table 9-1. Average DEFF and DEFT for estimates from the adult interview (continued)

Stratum	Design effect (*DEFF*) Average	Median	Maximum	Minimum	*DEFT* Average
San Bernardino	3.40	3.29	5.73	1.99	1.83
Riverside	3.47	3.05	6.29	1.16	1.83
Alameda	2.96	2.89	5.81	1.03	1.69
Sacramento	3.09	2.98	5.69	0.38	1.72
Contra Costa	2.83	2.91	5.13	0.12	1.65
Fresno	3.09	2.88	5.12	1.10	1.73
San Francisco	3.20	3.08	7.07	0.17	1.74
Ventura	2.38	2.33	4.50	0.04	1.50
San Mateo	2.79	2.89	5.31	0.22	1.61
Kern	3.43	3.42	6.51	0.37	1.82
San Joaquin	2.36	2.48	4.12	0.21	1.49
Sonoma	2.96	3.25	5.36	0.16	1.66
Stanislaus	3.44	3.41	8.78	0.14	1.80
Santa Barbara	2.87	2.61	6.26	0.14	1.64
Solano	2.96	2.98	6.41	0.88	1.69
Tulare	4.00	3.63	8.43	0.08	1.91
Santa Cruz	2.65	2.73	6.62	0.10	1.56
Marin	3.33	3.66	6.06	0.29	1.77
San Luis Obispo	2.78	2.85	6.97	0.17	1.58
Placer	3.13	3.00	5.62	0.65	1.73
Merced	3.51	3.47	9.49	0.14	1.81
Butte	2.25	2.34	4.13	0.27	1.46
Shasta	2.67	2.70	4.33	0.54	1.60
Yolo	4.24	3.13	11.41	0.07	1.90
EL Dorado	2.77	3.00	6.24	0.43	1.60
Imperial	3.45	3.52	9.89	0.67	1.79
Napa	4.14	3.62	9.72	1.35	1.98
Kings	4.03	4.11	7.80	0.04	1.95
Madera	2.83	2.66	10.13	0.81	1.64
Monterey	2.37	2.51	3.93	0.52	1.52
Humboldt	3.25	2.92	13.03	0.13	1.71
Nevada	3.82	3.75	7.83	0.16	1.90
Mendocino	3.62	3.22	8.48	0.65	1.85
Sutter	3.13	3.12	11.11	0.67	1.71
Yuba	2.90	2.86	8.13	0.07	1.62
Lake	3.57	3.25	8.43	0.16	1.79
San Benito	3.89	3.83	8.18	0.41	1.90
Colusa, Glen, Tehama	2.54	2.52	6.54	0.35	1.55
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	4.70	4.94	10.66	0.16	2.05
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	4.44	4.59	10.26	0.25	2.03

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

Table 9-2 shows the average *DEFT* for estimates from the child interview in each stratum for the combined landline/list/cell sample. The average *DEFT* at the state level is 2.07. In approximately 75 percent of the strata, the average *DEFT*s vary between 1.36 and 1.73; that is, the standard errors of these estimates are between 36 and 73 percent greater than expected from a simple random sample. As with the previous CHIS cycles, the state average *DEFT*s for estimates from the child interview are larger than those for the adult interview.

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

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
State	4.36	4.31	8.10	2.60	2.07
Los Angeles	3.63	3.57	5.68	2.03	1.89
San Diego	3.24	3.19	4.55	1.64	1.79
Orange	3.29	3.13	6.04	0.69	1.78
Santa Clara	3.41	3.29	9.14	0.53	1.76
San Bernardino	3.56	3.68	5.26	1.18	1.87
Riverside	2.80	3.14	6.91	0.35	1.59
Alameda	3.54	3.65	8.88	0.47	1.81
Sacramento	3.05	2.73	6.15	0.23	1.68
Contra Costa	2.69	3.10	6.01	0.28	1.53
Fresno	3.97	3.83	8.12	0.29	1.86
San Francisco	2.40	1.77	9.14	0.24	1.38
Ventura	1.78	1.83	3.39	0.47	1.30
San Mateo	2.45	2.50	5.89	0.12	1.43
Kern	1.96	1.87	3.61	0.04	1.33
San Joaquin	1.88	1.66	6.65	0.16	1.28
Sonoma	2.30	2.12	5.39	0.51	1.47
Stanislaus	2.17	2.35	3.81	0.08	1.40
Santa Barbara	2.68	2.19	10.71	0.15	1.47
Solano	2.15	2.20	3.58	0.54	1.42
Tulare	3.45	2.56	8.49	0.15	1.65
Santa Cruz	2.70	2.61	5.35	0.44	1.59
Marin	1.86	1.85	5.60	0.47	1.31
San Luis Obispo	2.25	1.85	6.13	0.18	1.41
Placer	1.45	1.42	2.87	0.18	1.13
Merced	3.30	3.11	7.16	0.20	1.70
Butte	2.40	2.46	6.74	0.26	1.44
Shasta	2.42	2.41	4.81	0.28	1.50
Yolo	2.68	2.02	10.88	0.25	1.43
EL Dorado	1.85	1.70	3.58	0.50	1.32
Imperial	4.08	4.11	9.05	0.40	1.94
Napa	3.10	3.10	6.37	0.24	1.66
Kings	2.95	3.32	5.90	0.33	1.61
Madera	3.08	3.29	6.52	0.35	1.69
Monterey	1.63	1.71	2.78	0.16	1.23
Humboldt	2.09	2.12	4.69	0.17	1.38
Nevada	1.69	1.69	3.22	0.17	1.25
Mendocino	1.14	1.26	2.03	0.18	1.02

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

Stratum	Design effect (*DEFF*) Average	Median	Maximum	Minimum	*DEFT* Average
Sutter	3.35	2.72	7.80	0.20	1.73
Yuba	2.55	2.42	5.77	0.28	1.48
Lake	1.71	1.82	3.08	0.24	1.25
San Benito	5.11	5.68	11.38	0.05	2.13
Colusa, Glen, Tehama	1.84	1.98	3.12	0.36	1.33
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	2.08	1.94	4.75	0.33	1.36
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	3.22	2.65	9.57	0.92	1.73

Source: UCLA Center for Health Policy Research, 2013-2014 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 1.73. In 33 strata (75 percent) the average *DEFT*s are between 1.18 and 1.50.

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

Stratum	Design effect (*DEFF*) Average	Median	Maximum	Minimum	*DEFT* Average
State	3.00	2.94	4.25	2.06	1.73
Los Angeles	2.98	3.05	4.26	1.64	1.71
San Diego	2.49	2.55	4.15	0.96	1.55
Orange	2.26	2.25	3.55	1.49	1.50
Santa Clara	2.01	1.87	3.77	0.29	1.37
San Bernardino	1.99	2.14	3.71	0.45	1.35
Riverside	2.61	2.60	5.43	0.56	1.58
Alameda	2.96	3.04	7.36	0.61	1.64
Sacramento	2.82	2.68	5.93	0.48	1.61
Contra Costa	2.14	2.39	3.41	0.92	1.45
Fresno	3.84	3.78	6.46	1.95	1.93
San Francisco	2.51	2.57	4.53	0.46	1.55
Ventura	1.48	1.25	3.09	0.43	1.18
San Mateo	1.56	1.60	4.37	0.46	1.19
Kern	1.97	1.66	3.92	0.15	1.35
San Joaquin	1.29	1.24	2.99	0.16	1.09
Sonoma	1.63	1.49	2.25	1.17	1.27
Stanislaus	1.03	1.00	1.78	0.37	1.00
Santa Barbara	1.46	1.55	3.15	0.09	1.13

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

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Solano	2.26	2.64	5.84	0.09	1.36
Tulare	2.19	2.15	2.87	1.53	1.48
Santa Cruz	0.93	0.97	1.26	0.42	0.96
Marin	1.21	1.34	2.12	0.36	1.08
San Luis Obispo	1.22	1.23	2.19	0.21	1.08
Placer	1.82	2.00	2.88	0.45	1.32
Merced	2.52	2.38	6.57	0.15	1.51
Butte	1.12	1.15	1.69	0.36	1.04
Shasta	1.22	1.05	2.37	0.26	1.06
Yolo	1.70	1.76	3.00	0.46	1.28
EL Dorado	1.20	1.33	1.76	0.28	1.08
Imperial	1.85	1.84	4.31	0.37	1.33
Napa	2.49	2.46	4.87	0.40	1.50
Kings	1.91	1.85	3.40	0.53	1.35
Madera	2.42	1.51	6.85	0.30	1.42
Monterey	1.75	1.35	4.17	0.18	1.22
Humboldt	1.46	1.45	2.78	0.72	1.19
Nevada	1.97	1.99	4.10	0.57	1.35
Mendocino	1.26	1.32	1.83	0.42	1.10
Sutter	1.48	1.49	2.53	0.40	1.18
Yuba	2.51	2.57	3.54	0.74	1.56
Lake	1.44	1.22	2.79	0.52	1.17
San Benito	1.54	1.55	2.32	0.90	1.23
Colusa, Glen, Tehama	2.07	2.18	4.75	0.40	1.39
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	1.76	1.42	4.48	0.19	1.26
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.56	1.63	2.93	0.69	1.23

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

9.2. Methods for Variance Estimation

Variance estimation procedures have been developed to account for a 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 used for variance computation are replication and linearization (i.e., the Taylor series approximation). Wolter (1985) presents a useful reference on the theory and applications of these methods. Shao (1996) has 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 subsample, 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 & 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:

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

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 2013-2014, a paired unit jackknife method (JK2[14]), a form of delete-a-group 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 2013-2014 are (1) operational convenience and (2) 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 2013-2014 analysis weights all affect the sampling errors of the estimates produced from the survey. Furthermore, the set of weights created in CHIS 2013-2014 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 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 2007 National Household Education Survey (Hagedorn, Roth, O'Donnell, Smith, & Mulligan, 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

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

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 2013-2014 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 & Bowles, 1996). In this section, we describe the elements needed to compute estimates for CHIS 2013-2014 using some of these programs.

WesVar Version 5.1 (Westat, 2007) 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 2013-2014 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, 2012) 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 complex survey data. Version 9.4 (SAS Institute Inc., 2015). In Version 9.4, 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 and 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 15 (StataCorp, 2015). 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.

Another software package that can be used to analyze survey data is R 3.2.0 (R Development Core Team, 2015) with the package *survey* (Lumley, 2012). R is a free software open source environment for statistical computing and graphics. It compiles and runs on UNIX platforms, Windows and MacOS. It is a command driven, fully programmable statistical language and environment used for managing, analyzing, and graphing data. The package *survey* has commands to analyze data from sample surveys such as description summary statistics, generalized linear models, Cox models, log-linear models, and general maximum pseudo-likelihood estimation for multistage stratified, cluster-sampled, unequally weighted survey samples. Like SUDAAN and SAS, the package *survey* in R can use linearization or replication depending how the design is defined using the command *svydesign*.

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 a combined data file with adults and teens, adults and children, or teens and children in the same file. 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. The variable TSVARSTR was created by sequentially numbering the sampling strata separately by sample type and year (i.e., landline sample, cell phone sample, list sample, and ABS sample).

- **TSVRUNIT** (Taylor's series unit). The variable TSVRUNIT indicates the PSU this case is the sampled household. In CHIS 2013-2014, the value of TSVARUNIT corresponds to the sequential numbering of sampled household within sampling stratum.

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

REFERENCES

Blumberg, S., & Luke., J. (2012, September). *Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, July-December 2012.* Retrieved October 3, 2013, from http://www.cdc.gov/nchs/nhis/releases.htm#wireless

Blumberg, S., Luke, J., & Cynamon, M. L. (2006). Telephone coverage and health survey estimates: Evaluating the need for concern about wireless substitution. *American Journal of Public Health, 95*(5), 926-931.

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

Brick, J. M., Flores Cervantes, I., Lee, S., & Norman, G. (2011, June). Nonsampling errors in dual frame telephone surveys. *Survey Methodology, 37*(1), 1-12.

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

Bureau, U. C. (n.d.). 2010 Census Summary File 2: California. Retrieved from http://www.census.gov/2010census/news/press-kits/sf2/summary-file-2.html

Bureau, U. C. (2012). 2010 Summary File 2: California. Retrieved from http://www.census.gov/2010census/news/press-kits/sf2/summary-file-2.html

Council of American Survey Research Organizations. (1982). *Special Report: On the Definition of Response Rates.* Port Jefferson, NY: CASRO.

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

Ford, B. M. (1983). An overview of hot-deck procedures. In W. Madow, N. H., & I. Olkin, *Incomplete data in sample surveys* (Vol. 2). New York: Academic Press.

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

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

Hartley, H. O. (1962). Multiple Frame Surveys. *Proceedings of the Social Statistics Section American Statistical Association,* (pp. 203-206).

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

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

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

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

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

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

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

Lumley, T. (2012). survey: analysis of complex survey samples. R package version 3.28-2. .

Norman, G., & Sigman, R. (2009). Using Addresses As Aampling Units in the 2007 Health Information National Trends Survey. *Joint Statistical Meetings. Proceedings of the Social Statistics Section American Statistical Association.* American Statistical Association.

Norman, G., & Sigman, R. (2009). Using Addresses as Sampling Units in the 2007 Health Information National Trends Survey. *Joint Statistical Meetings. Section on Survey Research Methods.*

Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. Retrieved from http://www.whitehouse.gov/omb/fedreg_1997standards

Oh, H., & Scheuren, F. (1983). Weighting adjustments for unit nonresponse. In W. Madow, I. Olkin, & D. Rubin, *Incomplete data in sample surveys* (Vol. 2, pp. 143-184). Academic Press.

R Development Core Team. (2015). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria. doi:http://www.R-project.org

Research Triangle Institute. (2012). *SUDAAN Language Manual, Volumes 1 and 2, Release 11*. Research Triangle Park, NC: Research Triangle Institute.

Rizzo, L., Brick, J., & Park, I. (2004). A Minimally Intrusive Method for Sampling Persons in Random Digit Dialing Surveys. *Public Opinion Quarterly, 68*, 267-274.

Rust, K., & Rao, J. (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. In W. Madow, H. Nisselson, & I. Olkin, *Incomplete data in sample surveys* (Vol. 3). New York: Academic Press.

SAS Institute Inc. (2015). *SAS/STAT® 14.1 User's Guide.* Cary, NC.

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

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

StataCorp. (2015). *Stata Statistical Software: Release 14.* College Station, TX: StataCorp LP.

State of California, Department of Finance. (2013). *Report P-1 (County, Race, Age): State and County Total Population Projections, 2010-2060.* Sacramento, California.

State of California, Department of Finance. (2014). *California County Population Estimates and Components of Change by Year, July 1, 2010–2014.* Sacramento, CA.

U.S. Census Bureau. (2012). 2010 Summary File 1: California. Retrieved from http://www.census.gov/2010census/news/press-kits/summary-file-1.html

U.S. Census Bureau. (2012). 2010 Summary File 2: California. Retrieved from http://www.census.gov/2010census/news/press-kits/sf2/summary-file-2.html

U.S. Census Bureau. (2012). Census 2010 Modified Race Data File. Retrieved from http://www.census.gov/popest/research/modified.html

U.S. Census Bureau. (2014, October 23). 2011-2013 3-year American Community Survey PUMS file. Retrieved from https://www.census.gov/programs-surveys/acs/data/pums.html

Verma, V., Scott, C., & 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. (2007). *WesVar ™ 4.0 user's guide.* Rockville, MD: Westat.

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

APPENDIX A

Table A-1. CHIS 2013-2014 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, and Japanese lists)

Sampling Stratum	Description	Landline sampling frame			Korean surname list			Korean & any other race but Vietnamese surname list			Vietnamese surname list		
		Average frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight
1.12	Los Angeles, San Fernando SPA – High Density	140,197	16,303	8.6	674	44	5.4	826	25	7.6	124	5	4.8
1.13	Los Angeles, San Gabriel SPA – High Density	365,881	26,307	13.8	1,168	69	8.5	7,492	193	11.6	3,417	77	13.5
1.14	Los Angeles, Metro SPA – High Density	397,762	17,789	22.3	1,821	100	10.7	1,935	55	15.6	185	4	18.5
1.17	Los Angeles, South SPA – High Density	114,672	14,536	7.9	549	29	6.0	1,106	30	6.7	239	3	8.9
1.18	Los Angeles, South Bay SPA – High Density	207,916	12,954	16.0	432	26	8.3	740	14	12.8	489	10	11.9
1.21	Los Angeles, Antelope Valley SPA – Low Density	229,115	9,637	23.8	131	8	11.9	113	4	14.1	156	1	22.3
1.22	Los Angeles, San Fernando SPA – Low Density	1,715,917	19,736	86.9	1,557	83	15.9	2,148	54	29.8	1,596	36	28.0
1.23	Los Angeles, San Gabriel SPA – Low Density	949,492	21,639	43.7	1,091	56	13.8	6,343	178	21.3	3,651	84	26.6
1.24	Los Angeles, Metro SPA – Low Density	629,281	17,169	36.6	791	46	11.6	1,681	43	20.5	699	16	20.6
1.25	Los Angeles, West SPA – Low Density	1,009,426	17,669	57.1	631	36	15.0	1,237	33	25.2	309	7	22.1
1.26	Los Angeles, South SPA –Low Density	688,945	19,509	35.3	548	33	12.2	547	15	20.3	219	7	18.3
1.27	Los Angeles, East SPA – Low Density	779,483	17,112	45.5	608	34	14.1	1,109	28	22.2	488	11	23.2
1.28	Los Angeles, South Bay SPA – Low Density	1,069,775	13,850	77.3	1,319	69	14.7	2,039	63	22.2	1,350	33	28.7
2.12	San Diego North Central SR-High	212,143	11,947	17.7	159	8	12.2	332	8	15.8	1,527	36	14.8
2.13	San Diego Central SR- High	184,694	10,205	18.1	96	7	10.7	140	4	10.0	1,025	24	14.4
2.21	San Diego North Coastal SR-Low	357,952	16,144	22.2	226	11	8.7	303	7	15.2	378	6	15.8
2.22	San Diego North Central SR -Low	354,487	11,603	30.5	248	14	10.3	562	12	22.5	460	8	20.0
2.23	San Diego Central SR - Low	296,833	18,555	16.0	81	4	7.4	155	3	9.7	303	10	11.7
2.24	San Diego South SR- Low	345,366	24,313	14.2	231	8	11.6	280	8	13.3	232	5	12.9
2.25	San Diego East SR- Low	394,188	14,673	26.9	146	8	12.2	139	5	17.4	323	7	15.4

Table A-1. CHIS 2013-2014 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, and Japanese lists) (continued)

Sampling Stratum	Description	Landline sampling frame			Korean surname list			Korean & any other race but Vietnamese surname list			Vietnamese surname list		
		Average frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight
2.26	San Diego North Inland SR - Low	392,122	14,604	26.8	329	17	11.8	624	19	15.6	764	15	18.2
3.1	Orange – High Density	956,033	48,884	19.5	2,376	135	10.4	3,905	101	14.1	12,539	287	15.0
3.2	Orange – Low Density	1,722,009	32,231	53.4	1,601	88	13.9	3,077	80	24.0	4,746	100	26.7
4.1	Santa Clara - High	552,024	29,606	18.5	577	34	12.0	2,378	63	14.9	6,265	143	16.3
4.2	Santa Clara - Low	1,070,700	16,996	63.0	1,576	88	13.8	6,761	179	24.8	6,913	149	26.4
5	San Bernardino	1,330,056	35,727	37.2	1,036	54	12.3	1,970	54	20.1	1,678	37	20.7
6	Riverside	1,401,938	37,026	37.9	1,006	64	11.2	1,297	33	19.7	1,739	41	19.3
7	Alameda	1,533,689	33,973	45.1	1,832	107	12.6	8,921	243	21.4	4,528	100	24.0
8	Sacramento	1,188,904	28,852	41.2	1,032	61	11.7	2,854	71	23.2	4,058	95	21.6
9	Contra Costa	970,239	21,743	44.6	808	38	14.2	2,473	64	23.3	1,177	29	23.1
10	Fresno	649,200	16,600	39.1	435	25	15.0	791	20	22.0	637	14	27.7
11	San Francisco	1,079,576	27,612	39.0	1,476	87	12.7	11,005	293	21.7	2,696	63	22.8
12	Ventura	637,892	16,581	38.5	376	21	13.0	662	20	17.9	610	15	18.5
13	San Mateo	813,551	17,527	46.4	905	46	14.8	4,042	108	21.2	678	13	24.2
14	Kern	521,921	10,744	48.6	236	14	14.8	181	4	18.1	267	6	22.3
15	San Joaquin	435,847	11,362	38.3	325	14	19.1	616	15	22.8	1,005	24	20.9
16	Sonoma	457,724	9,592	47.7	288	21	9.9	296	9	24.7	387	9	32.3

Table A-1. CHIS 2013-2014 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, and Japanese lists) (continued)

		Landline sampling frame			Korean surname list			Korean & any other race but Vietnamese surname list			Vietnamese surname list		
Sampling Stratum	Description	Average frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight
17	Stanislaus	334,836	12,473	26.9	208	11	12.2	208	5	11.6	274	9	14.4
18	Santa Barbara	376,387	11,900	31.6	175	10	13.5	199	5	22.1	197	5	21.9
19	Solano	319,371	12,901	24.8	166	10	9.8	318	10	21.2	336	6	16.0
20	Tulare	247,676	10,636	23.3	79	5	7.9	110	5	7.9	107	2	26.8
21	Santa Cruz	266,540	11,301	23.6	123	9	10.3	171	3	28.5	117	2	14.6
22	Marin	332,046	11,922	27.8	198	12	9.0	339	11	17.8	291	5	26.5
23	San Luis Obispo	243,269	9,615	25.3	92	6	13.1	86	1	17.2	111	3	13.9
24	Placer	297,637	11,164	26.7	204	7	17.0	307	10	15.4	305	6	25.4
25	Merced	128,851	11,098	11.6	83	6	5.9	108	6	7.2	98	1	14.0
26	Butte	168,995	6,270	26.9	103	6	12.9	104	3	20.8	166	1	55.3
27	Shasta	154,720	6,262	24.7	97	4	8.8	44	1	11.0	79	3	15.8
28	Yolo	144,368	10,104	14.3	93	6	10.3	285	7	16.8	124	4	17.7
29	El Dorado	175,490	9,796	17.9	93	6	9.3	123	3	15.4	104	1	20.8
30	Imperial	89,437	8,640	10.3	33	2	16.5	54	1	6.8	26	1	8.7
31	Napa	116,963	12,080	9.7	52	2	6.5	54	1	9.0	52	1	8.7
32	Kings	71,732	13,277	5.4	38	1	9.5	38	0	0.0	31	0	0.0
33	Madera	87,228	10,413	8.4	53	2	4.4	25	1	8.3	36	0	0.0

Table A-1. CHIS 2013-2014 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, and Japanese lists) (continued)

		Landline sampling frame			Korean surname list			Korean & any other race but Vietnamese surname list			Vietnamese surname list		
Sampling Stratum	Description	Average frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight
34	Monterey	350,301	14,209	24.6	209	12	12.3	277	7	17.3	243	6	16.2
35	Humboldt	118,827	6,934	17.1	55	4	11.0	41	3	10.3	52	1	10.4
36	Nevada	99,162	8,583	11.5	78	4	8.7	72	2	10.3	58	3	9.7
37	Mendocino	80,081	7,921	10.1	36	2	6.0	42	0	0.0	53	1	13.3
38	Sutter	55,147	9,526	5.8	39	3	4.9	37	1	6.2	47	1	5.2
39	Yuba	51,981	10,413	5.0	41	0	0.0	64	3	4.6	41	2	8.2
40	Lake	65,109	8,077	8.1	34	1	34.0	27	0	0.0	17	2	5.7
41	San Benito	36,675	17,266	2.1	15	0	0.0	6	1	2.0	8	0	0.0
42	Colusa, Glenn, Tehama	79,486	6,941	11.5	34	3	6.8	19	2	4.8	38	0	0.0
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Trinity	104,097	4,826	21.5	46	2	23.0	31	0	0.0	72	2	36.0
43.2	Siskiyou	58,460	12,282	4.8	21	1	10.5	14	0	0.0	20	0	0.0
44	Alpine, Amador, Inyo, Mariposa, Mono	114,372	2,387	47.8	47	3	15.7	37	0	0.0	97	1	48.5
44.2	Tuolumne	55,551	6,317	8.8	22	2	11.0	15	0	0.0	26	0	0.0
44.3	Calaveras	63,922	10,747	5.9	28	1	5.6	20	0	0.0	28	1	7.0
	Total	31,045,668	1,047,591		31,316	1,750		84,355	2,250		71,111	1,600	

Table A-1. CHIS 2013-2014 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, and Japanese lists) (continued)

		Japanese surname and non-Japanese first name			Japanese first name and non-Japanese surname			Japanese surname and first name		
Sampling Stratum	Description	Frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight
1.12	Los Angeles, San Fernando SPA – High Density	269	10	6.0	263	8	7.1	107	7	4.9
1.13	Los Angeles, San Gabriel SPA – High Density	916	40	9.3	527	16	10.5	423	49	5.8
1.14	Los Angeles, Metro SPA – High Density	313	11	14.9	255	8	10.6	173	23	5.4
1.17	Los Angeles, South SPA – High Density	319	12	7.3	128	4	6.7	110	9	6.1
1.18	Los Angeles, South Bay SPA – High Density	1,599	64	11.0	339	10	10.9	953	93	6.4
1.21	Los Angeles, Antelope Valley SPA – Low Density	159	6	12.2	203	6	10.7	14	3	2.8
1.22	Los Angeles, San Fernando SPA – Low Density	1,600	65	20.5	1,871	55	25.3	478	54	8.7
1.23	Los Angeles, San Gabriel SPA – Low Density	1,298	55	15.8	1,122	31	21.2	430	41	8.8
1.24	Los Angeles, Metro SPA – Low Density	633	29	12.9	582	17	20.1	256	24	8.8
1.25	Los Angeles, West SPA – Low Density	1,190	47	16.8	748	17	31.2	727	81	7.9
1.26	Los Angeles, South SPA –Low Density	650	30	14.8	528	14	19.6	137	14	7.6
1.27	Los Angeles, East SPA – Low Density	1,339	56	15.9	740	19	20.0	386	36	9.0
1.28	Los Angeles, South Bay SPA – Low Density	2,147	98	16.5	1,200	41	23.1	1,102	123	8.0
2.12	San Diego North Central SR-High	163	7	14.8	234	8	10.6	62	7	7.8
2.13	San Diego Central SR- High	188	9	11.1	186	5	11.6	24	0	0.0
2.21	San Diego North Coastal SR-Low	377	12	12.2	501	16	14.7	91	9	8.3
2.22	San Diego North Central SR -Low	212	11	13.3	294	8	19.6	109	12	6.8
2.23	San Diego Central SR - Low	179	6	11.2	184	4	15.3	21	5	4.2
2.24	San Diego South SR- Low	481	21	9.6	438	13	8.6	107	10	5.9
2.25	San Diego East SR- Low	340	17	11.7	371	11	13.7	22	4	5.5

Table A-1. CHIS 2013-2014 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, and Japanese lists) (continued)

Sampling Stratum	Description	Japanese surname and non-Japanese first name			Japanese first name and non-Japanese surname			Japanese surname and first name		
		Frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight
2.26	San Diego North Inland SR - Low	333	14	14.5	489	11	16.3	91	9	8.3
3.1	Orange – High Density	1,436	57	10.3	818	21	10.2	685	80	6.2
3.2	Orange – Low Density	2,203	92	16.4	1,661	45	19.1	713	72	8.0
4.1	Santa Clara - High	826	30	13.3	441	15	11.6	221	27	6.1
4.2	Santa Clara - Low	2,511	108	17.2	1,413	36	24.8	938	94	8.0
5	San Bernardino	1,235	53	14.4	1,304	40	16.9	130	15	6.5
6	Riverside	1,309	58	13.4	1,665	49	17.3	170	18	7.1
7	Alameda	2,112	84	16.1	1,806	49	22.6	595	66	6.8
8	Sacramento	1,862	78	15.5	1,320	37	16.1	360	41	7.7
9	Contra Costa	1,397	58	14.7	1,128	33	17.6	274	29	8.6
10	Fresno	1,115	46	15.1	584	17	16.2	215	25	7.2
11	San Francisco	1,136	49	14.6	1,473	44	17.1	566	59	7.9
12	Ventura	788	35	14.1	691	21	19.7	180	18	8.6
13	San Mateo	1,203	53	16.7	1,011	26	22.5	552	62	7.7
14	Kern	346	14	15.0	405	10	23.8	25	1	12.5
15	San Joaquin	717	30	13.5	477	17	15.4	135	15	7.9
16	Sonoma	423	18	16.9	627	20	19.0	67	9	6.7

Table A-1. CHIS 2013-2014 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, and Japanese lists) (continued)

		Japanese surname and non-Japanese first name			Japanese first name and non-Japanese surname			Japanese surname and first name		
Sampling Stratum	Description	Frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight
17	Stanislaus	259	9	11.3	331	8	19.5	24	2	8.0
18	Santa Barbara	403	17	12.2	340	12	16.2	63	7	7.0
19	Solano	303	14	12.1	512	17	13.1	38	2	5.4
20	Tulare	242	9	13.4	223	8	17.2	23	2	7.7
21	Santa Cruz	371	14	14.3	255	7	14.2	88	11	6.3
22	Marin	289	13	10.7	390	9	21.7	96	10	6.9
23	San Luis Obispo	220	11	12.9	239	6	12.6	35	5	7.0
24	Placer	391	16	13.0	379	10	12.2	63	5	9.0
25	Merced	130	5	8.7	122	3	7.6	10	3	2.5
26	Butte	82	5	9.1	169	7	11.3	8	0	0.0
27	Shasta	77	3	11.0	160	4	10.7	7	2	3.5
28	Yolo	219	9	8.8	166	6	7.2	51	5	7.3
29	El Dorado	179	5	11.2	202	6	11.2	22	3	5.5
30	Imperial	152	4	8.0	105	3	13.1	2	0	0.0
31	Napa	93	3	10.3	131	5	8.7	9	1	4.5
32	Kings	86	4	4.3	79	1	3.6	7	0	0.0
33	Madera	66	4	5.1	101	2	6.7	10	1	10.0

Table A-1. CHIS 2013-2014 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, and Japanese lists) (continued)

		Japanese surname and non-Japanese first name			Japanese first name and non-Japanese surname			Japanese surname and first name		
Sampling Stratum	Description	Frame size	Sample size	Weight	Frame size	Sample size	Weight	Frame size	Sample size	Weight
34	Monterey	455	18	11.7	375	10	16.3	128	12	7.5
35	Humboldt	67	2	8.4	133	4	8.3	6	0	0.0
36	Nevada	103	2	11.4	142	2	8.9	8	1	4.0
37	Mendocino	39	1	7.8	93	3	7.8	5	0	0.0
38	Sutter	98	4	6.1	57	2	5.7	15	0	0.0
39	Yuba	54	0	0.0	54	1	13.5	2	0	0.0
40	Lake	25	1	3.6	60	1	12.0	0	0	0.0
41	San Benito	37	1	2.1	28	1	2.3	4	0	0.0
42	Colusa, Glenn, Tehama	34	0	0.0	56	2	7.0	4	1	1.3
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Trinity	39	4	9.8	103	1	20.6	0	0	0.0
43.2	Siskiyou	11	0	0.0	62	2	3.9	2	0	0.0
44	Alpine, Amador, Inyo, Mariposa, Mono	99	6	12.4	117	5	23.4	13	0	0.0
44.2	Tuolumne	31	0	0.0	54	0	0.0	2	0	0.0
44.3	Calaveras	28	0	0.0	59	0	0.0	2	0	0.0
	Total	40,006	1,667		33,324	950		12,396	1,317	

[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 2013-2014 cell-phone sample frame size, sample sizes, and base weights by sampling stratum or area code

Sampling stratum	Description	Cell phone sampling frame		
		Average frame size	Sample size	Weight
1	Los Angeles	19,036	13,433,652	705.7
2	San Diego	12,575	4,138,739	329.1
3	Orange	5,550	4,142,114	746.3
4	Santa Clara	3,991	2,393,046	599.6
5	San Bernardino	3,251	2,451,596	754.1
6	Riverside	2,939	2,384,872	811.5
7	Alameda	3,725	2,286,138	613.7
8	Sacramento	2,177	1,712,154	786.5
9	Contra Costa	1,586	1,005,925	634.3
10	Fresno	1,491	1,151,081	772.0
11	San Francisco	2,502	1,502,340	600.5
12	Ventura	1,563	1,004,011	642.4
13	San Mateo	1,157	752,244	650.2
14	Kern	1,073	982,805	915.9
15	San Joaquin	1,381	741,557	537.0
16	Sonoma	1,440	582,315	404.4
17	Stanislaus	1,170	598,113	511.2
18	Santa Barbara	831	520,674	626.6
19	Solano	1,312	464,254	353.9
20	Tulare	985	419,209	425.6
21	Santa Cruz	1,135	289,778	255.3
22	Marin	1,438	296,095	205.9
23	San Luis Obispo	1,006	317,995	316.1
24	Placer	1,254	402,895	321.3
25	Merced	1,348	236,423	175.4
26	Butte	914	254,343	278.3
27	Shasta	1,369	218,002	159.2
28	Yolo	1,503	160,846	107.0
29	El Dorado	1,177	140,258	119.2
30	Imperial	2,595	258,913	99.8
31	Napa	1,495	118,101	79.0
32	Kings	1,166	137,477	117.9
33	Madera	1,652	132,526	80.2
34	Monterey	1,116	501,583	449.4
35	Humboldt	996	155,520	156.1
36	Nevada	1,105	94,656	85.7
37	Mendocino	1,645	111,545	67.8
38	Sutter	3,668	196,034	53.4
40	Lake	1,186	53,203	44.9
41	San Benito	1,483	64,400	43.4
42	Colusa, Glenn, Tehama	876	69,705	79.6
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	733	151,179	206.2

Table A-2. CHIS 2013-2014 cell-phone sample frame size, sample sizes, and base weights by sampling stratum or area code (continued)

Sampling stratum	Description	Cell phone sampling frame		
		Average frame size	Sample size	Weight
44	Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1,335	159,839	119.7
	Total	101,930	47,188,156	

Table A-3. CHIS 2013-2014 sample frame size, sample size, and base weight for the ABS sample

Sampling stratum	Description	ABS		
		Frame size	Sample size	Weight
16	Sonoma	191,602	5,394	35.5

APPENDIX B

Table B-1. Screener interview (households) weighting adjustments by sample type

	Landline	List	Cell	Sonoma ABS
1. Base weight				
1.1 Sample size	1,047,591	9,534	101,930	5,394
1.2 Sum of weights	30,896,715	148,953	47,188,156	191,602
1.3 Coefficient of variation	59.93	38.48	56.90	0
2. Adjusting for unreleased landline cell phone cases				
2.1 Sum of weights before adjustment				
a. No adjustment	30,593,545	148,953	37,918,327	191,602
b. Cell inactive/unknown - fielded	0	0	5,733,204	NA
c. Cell inactive/unknown - not fielded	0	0	3,839,794	NA
2.2 Sum of weights after adjustment				
a. No adjustment	30,593,545	148,953	37,918,327	191,602
b. Cell inactive/unknown - fielded	0	0	9,572,998	NA
c. Cell inactive/unknown - not fielded	0	0	0	NA
2.3 Sample size	763,994	26,248	137,525	5,394
2.4 Coefficient of variation	59.90	38.48	72.58	0
3. CATI extraction and adjusting for new work subsampling				
3.1 Sum of weights before adjustment				
a. Purged (nonresidential) telephone number	19,965,832	29,573	0	NA
a. Not purged, not subsampled	10,453,106	119,381	47,374,417	191,602
b. Not purged, subsampled	174,608	0	116,908	NA
3.2 Sum of weights after adjustment				
a. Purged (nonresidential) telephone number	19,965,832	29,573	0	NA
a. Not purged, not subsampled	10,627,714	119,381	47,491,325	191,602
b. Not purged, subsampled	0	0	0	NA
3.3 Sample size	745,866	26,248	124,156	5,394
3.4 Coefficient of variation	59.79	38.48	72.40	0
4. First refusal conversion subsampling adjustment				
4.1 Sum of weights before adjustment				
a. Household never refused	7,644,961	72,914	29,958,422	191,602
b. Household refused - selected for refusal conversion	2,982,753	46,466	17,532,904	NA
c. Household refused - not selected for refusal conversion	0	0	0	NA

Table B-1. Screener interview (households) weighting adjustments by sample type (continued)

	Landline	List	Cell	Sonoma ABS
4.2 Sum of weights after adjustment				
a. Household never refused	7,644,961	72,914	29,958,422	191,602
b. Household refused - selected for refusal conversion	2,982,753	46,466	17,532,904	NA
c. Household refused - not selected for refusal conversion	0	0	0	NA
4.3 Sample size	322,796	18,700	124,156	5,394
4.4 Coefficient of variation	59.43	38.30	72.40	0
5. Second refusal conversion subsampling				
5.1 Sum of weights before adjustment				
a. Household never refused more than once	8,326,782	83,449	33,694,544	191,602
b. Household refused -selected for second refusal conversion	2,300,408	35,058	13,795,149	NA
c. Household refused -not selected for second refusal conversion	523	874	1,633	NA
5.2 Sum of weights after adjustment				
a. Household never refused more than once	8,326,782	83,449	33,694,544	191,602
b. Household refused -selected for second refusal conversion	2,300,931	35,932	13,796,782	NA
c. Household refused -not selected for second refusal conversion	0	0	0	NA
5.3 Sample size				5,394
5.4 Coefficient of variation	59.43	37.46	72.39	0
6. Adjusting for unknown residential status				
6.1 Sum of weights before adjustment				
a. Residential - respondents	1,810,215	24,083	10,460,500	28,488
b. Residential - nonrespondents	2,350,287	44,183	13,928,724	157,146
c. Unknown residential status (NA, NM)	3,827,357	37,110	9,138,871	0
d. Nonresidential	2,639,855	14,005	13,963,229	5,968
6.2 Sum of weights - allocating unknown residential				
a. Residential - respondents	1,810,215	24,083	10,460,500	NA
b. Residential - nonrespondents	2,350,287	44,183	13,928,724	NA
c. (NA, NM)	2,102,082	31,494	5,900,733	NA
6.3 Sum of weights after adjustment				
a. Residential - respondents	1,810,215	24,083	10,460,500	28,488
b. Residential - nonrespondents	4,452,369	75,677	19,829,457	157,146
c. Estimated residential among unknown	0	0	0	NA
d. Nonresidential	2,639,855	14,005	13,963,229	5,968

Table B-1. Screener interview (households) weighting adjustments by sample type (continued)

	Landline	List	Cell	Sonoma ABS
6.4 Sample size	142,364	4,174	57,293	5,394
6.5 Coefficient of variation	71.80	43.83	74.21	0
7. Supplemental list-sample eligibility adjustment				
7.1 Sum of weights before adjustment				
a. RDD, Cell sample, ABS	6,262,583	0	30,289,958	191,602
b. Completed Korean, Vietnamese, or Japanese	0	11,734	0	NA
c. Nonresponse, but known that is not Korean, Vietnamese, or Japanese	0	12,592	0	NA
d. Nonresponse, unknown Korean, Vietnamese, or Japanese status	0	75,434	0	NA
7.2 Sum of weights after adjustment				191,602
a. RDD or Cell sample	6,262,583	0	30,289,958	NA
b. Completed Korean, Vietnamese, or Japanese	0	47,529	0	NA
c. Nonresponse, but known that is not Korean, Vietnamese, or Japanese	0	52,231	0	NA
d. Nonresponse, unknown Korean, Vietnamese, or Japanese status	0	0	0	NA
7.3 Sample size	142,364	1,529	57,293	5,394
7.4 Coefficient of variation	71.80	43.25	74.21	0
8. Unknown presence of children in household				
8.1 Sum of weights before adjustment				191,602
a. Ineligible respondent	9,728	0	4,259,888	NA
b. Eligible respondent - child status known	1,800,486	98,871	6,200,612	NA
c. Eligible nonrespondent - child status known	26,604	413	18,859	NA
d. Unknown nonrespondent - child status unknown	4,425,764	476	19,810,598	NA
8.2 Sum of weights after adjustment				191,602
a. Ineligible respondent	9,728	0	4,259,888	NA
b. Eligible respondent - child status known	1,827,091	99,284	6,219,471	NA
c. Eligible nonrespondent - child status known	0	0	0	NA
d. Unknown nonrespondent - child status unknown	4,425,764	476	19,810,598	NA
8.3 Sample Size	141,871	1,524	57,261	5,394
8.4 Coefficient of variation	71.46	42.92	74.19	0
9. Screener nonresponse adjustment				
9.1 Sum of weights before adjustment				
a. Respondents	1,836,819	47,053	10,479,359	28,488
b. Nonrespondents	4,425,764	52,706	19,810,598	157,146

Table B-1. Screener interview (households) weighting adjustments by sample type (continued)

	Landline	List	Cell	Sonoma ABS
9.2 Sum of weights after adjustment				
a. Respondents	6,262,583	99,760	30,289,958	185,634
b. Nonrespondents	0	0	0	0
9.3 Sample size	62,583	767	24,868	802
9.4 Coefficient of variation	71.47	117.64	73.57	18.4
10. Multiple telephone adjustment				
10.1 Sum of weights before adjustment	6,220,378	99,760	17,954,859	185,634
10.2 Sum of weights after adjustment	6,117,050	99,760	17,954,859	NA
10.3 Sample size	62,283	767	14,272	802
10.4 Coefficient of variation	71.87	117.64	70.12	18.4
10.5 Overall adjustment factor	98.3%	100.0%	100.0%	NA
11. Duplicate respondent adjustment				
11.1 Sum of weights before adjustment				
a. Not a duplicate number	6,116,070	99,760	17,948,116	185,634
b. Duplicate number	979	0	6,743	NA
11.2 Sum of weights after adjustment				
a. Not a duplicate number	6,117,050	99,760	17,954,859	185,634
b. Duplicate number	0	0	0	NA
11.3 Sample size	62,275	767	14,264	802
11.4 Coefficient of variation	71.87	117.64	70.10	18.4
12. Section G nonresponse adjustment*				
12.1 Sum of weights before adjustment				
a. Household with child 1st procedure	339,568	5,933	0	8,346
b. Household w/o child 1st procedure - section G completed	3,027,889	49,068	10,085,190	111,370
c. Household w/o child 1st procedure - section G not completed	2,749,593	44,759	7,869,669	65,919
12.2 Sum of weights after adjustment				
a. Household with child 1st procedure	339,568	5,933	0	8,346
b. Household w/o child 1st procedure - section G completed	5,777,482	93,827	17,954,859	177,289
c. Household w/o child 1st procedure - section G not completed	0	0	0	0
12.3 Sample size	34,864	447	8,064	516
12.4 Coefficient of variation	80.63	117.98	71.63	18.3

* 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

	Landline	List	Cell	Sonoma ABS
1. Adult Base Weight				
1.1 Number of Sampled Adults	62,275	767	14,264	802
1.2 Sum of Weights	13,028,113	205,679	20,214,178	185,634
1.3 Coefficient of Variation	92.94	125.04	89.11	18.85
2. Nonresponse Adjustment				
2.1 Sum of Weights Before Adjustment				
a. Eligible Respondents	5,795,663	94,425	10,463,541	191,142
b. Ineligible	86,463	2,942	113,506	1,044
c. Nonrespondents	7,145,987	108,312	9,637,131	158,364
2.2 Sum of Weights After Adjustment				
a. Eligible Respondents	12,753,712	200,149	19,964,190	348,085
b. Ineligible	274,401	5,530	249,988	2,463
c. Nonrespondents	0	0	0	0
2.3 Number of Completed Interviews	31,615	392	7,752	481
2.4 Coefficient of Variation (CV)	122.88	172.06	89.23	81.44
2.5 Mean Adjustment Factor	2.20	1.91	1.85	1.821

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

	Landline	List	Cell	Sonoma ABS
1. Child Base Weight				
1.1 Number of Sampled Children	5,978	69	1,701	64
1.2 Sum of Weights	2,122,502	20,323	7,619,569	14,822
1.3 Coefficient of Variation	131.49	113.85	108.25	18.07
2. Nonresponse Adjustment				
2.1 Sum of Weights Before Adjustment				
a. Eligible Respondents	1,448,612	15,778	5,475,298	22,664
b. Ineligible	17,173	0	33,330	388
c. Nonrespondents	656,717	4,545	2,110,941	10,053
2.2 Sum of Weights After Adjustment				
a. Eligible Respondents	2,097,662	20,323	7,571,525	32,437
b. Ineligible	24,839	0	48,044	669
c. Nonrespondents	0	0	0	0
2.3 Number of Completed Interviews	4,164	50	1,523	42
2.4 Coefficient of Variation (CV)	128.01	115.86	104.16	57.87
2.5 Mean Adjustment Factor	1.45	1.29	1.38	1.431

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

	Landline	List	Cell	Sonoma ABS
1. Teen Base Weight				
1.1 Number of Sampled Children	4,339	53	1,138	46
1.2 Sum of Weights	1,438,382	16,314	3,470,552	10,836
1.3 Coefficient of Variation	115.57	117.29	90.48	18.96
2. Nonresponse Adjustment				
2.1 Sum of Weights Before Adjustment				
a. Eligible Respondents	575,801	5,347	1,417,806	8,597
b. Ineligible	12,228	0	24,377	0
c. Nonrespondents	850,352	10,967	2,028,368	18,761
2.2 Sum of Weights After Adjustment				
a. Eligible Respondents	1,409,098	16,314	3,391,352	27,357
b. Ineligible	29,284	0	79,200	0
c. Nonrespondents	0	0	0	0
2.3 Number of Completed Interviews	1,738	18	482	15
2.4 Coefficient of Variation (CV)	113.32	81.21	96.54	60.19
2.5 Mean Adjustment Factor	2.45	3.05	2.39	3.182

Table B-5. Poststratification, composite weighting procedures, trimming, and raking adjustments for adult interviews by telephone and ABS stratum

	Landline/list/cell phone	Sonoma ABS
1. Poststratification to telephone service		
1.1 Number of Completed Interviews	39,759	481
1.2 Sum of weights before poststratification	32,918,050	348,085
1.3 Sum of weights after poststratification	42,147,907	348,085
2. Composite weight		
2.1 Sum of weights after composite factor	28,539,233	348,085
3. Trimming Adjustment*		
3.1 Number of Trimmed Records	47	2
3.2 Sum of Weights Before Trimming Adjustment	28,539,233	348,085
3.3 Sum of Weights After Trimming Adjustment	28,299,109	345,853
4. Raking Adjustment*		
4.1 Number of Completed Interviews	39,759	481
4.2 Sum of Weights After Adjustment	28,539,233	382,292
4.3 Coefficient of Variation (CV)	200.90	157.64
4.4 Mean Adjustment Factor	1.01	1.105
4.5 Mean Weight	717.81	794.79

*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 interviews by telephone and ABS stratum

	Landline/list/ cell phone	Sonoma ABS
1. Poststratification to telephone service		
1.1 Number of Completed Interviews	5,470	42
1.2 Sum of weights before poststratification	9,689,511	32,437
1.3 Sum of weights after poststratification	8,518,076	32,437
2. Composite weight		
2.1 Sum of weights after composite factor	6,134,953	32,437
3. Trimming Adjustment*		
3.1 Number of Trimmed Records	73	0
3.2 Sum of Weights Before Trimming Adjustment	6,134,953	32,437
3.3 Sum of Weights After Trimming Adjustment	5,777,641	32,437
4. Raking Adjustment*		
4.1 Number of Completed Interviews	5,470	42
4.2 Sum of Weights After Adjustment	6,006,452	58,705
4.3 Coefficient of Variation (CV)	197.86	74.25
4.4 Mean Adjustment Factor	1.04	1.810
4.5 Mean Weight	1,098.07	1,397.73

*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 telephone and ABS stratum

	Landline/list/ cell phone	Sonoma ABS
1. Poststratification to telephone service		
1.1 Number of Completed Interviews	2,238	15
1.2 Sum of weights before poststratification	4,816,764	27,357
1.3 Sum of weights after poststratification	4,520,752	27,357
2. Composite weight		
2.1 Sum of weights after composite factor	2,908,247	27,357
3. Trimming Adjustment*		
3.1 Number of Trimmed Records	76	0
3.2 Sum of Weights Before Trimming Adjustment	2,908,247	27,357
3.3 Sum of Weights After Trimming Adjustment	2,659,297	27,357
4. Raking Adjustment*		
4.1 Number of Completed Interviews	2,238	15
4.2 Sum of Weights After Adjustment	3,036,748	43,031
4.3 Coefficient of Variation (CV)	164.57	59.27
4.4 Mean Adjustment Factor	1.14	1.573
4.5 Mean Weight	1356.90	2,868.73

*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-8. Composite ABS-telephone sample, second trimming, and second raking adjustments for adult interviews

	All Samples
1. Composite ABS- telephone samples	
1.1 Number of Completed Interviews	40,240
1.2 Sum of weights before composite factor	28,921,524
1.3 Sum of weights after composite factor	28,539,233
1.4 Coefficient of Variation (CV)	201.72
2. Trimming Adjustment	
3.1 Number of Trimmed Records	47
3.2 Sum of Weights Before Trimming Adjustment	28,539,233
3.3 Sum of Weights After Trimming Adjustment	28,299,109
3. Raking Adjustment	
3.1 Number of Completed Interviews	39,759
3.2 Sum of Weights After Adjustment	28,539,233
3.3 Coefficient of Variation (CV)	200.9

Table B-9. Composite ABS-telephone sample, second trimming, and second raking adjustments for child interviews

	All Samples
1. Composite ABS- telephone samples	
1.1 Number of Completed Interviews	5,512
1.2 Sum of weights before composite factor	6,065,156
1.3 Sum of weights after composite factor	6,002,596
1.4 Coefficient of Variation (CV)	198.56
2. Trimming Adjustment	
3.1 Number of Trimmed Records	0
3.2 Sum of Weights Before Trimming Adjustment	6,002,596
3.3 Sum of Weights After Trimming Adjustment	6,002,596
3. Raking Adjustment	
3.1 Number of Completed Interviews	5,512
3.2 Sum of Weights After Adjustment	6,006,452
3.3 Coefficient of Variation (CV)	198.49

Table B-10. Composite ABS-telephone sample, second trimming, and second raking adjustments for adolescent interviews

	All Samples
1. Composite ABS- telephone samples	
1.1 Number of Completed Interviews	2,253
1.2 Sum of weights before composite factor	3,079,779
1.3 Sum of weights after composite factor	3,040,603
1.4 Coefficient of Variation (CV)	164.91
2. Trimming Adjustment	
3.1 Number of Trimmed Records	0
3.2 Sum of Weights Before Trimming Adjustment	3,040,603
3.3 Sum of Weights After Trimming Adjustment	3,040,603
3. Raking Adjustment	
3.1 Number of Completed Interviews	2,253
3.2 Sum of Weights After Adjustment	3,036,748
3.3 Coefficient of Variation (CV)	165.10