November 2017

CHIS 2015-2016 Methodology Report Series

Report 5

Weighting and Variance Estimation

CALIFORNIA HEALTH INTERVIEW SURVEY

CHIS 2015-2016 METHODOLOGY SERIES

REPORT 5

WEIGHTING AND VARIANCE ESTIMATION

NOVEMBER 2017

This report was prepared for the California Health Interview Survey by Jill A. Dever, Jamie L. Ridenhour, Michael A. Jacobsen, and Douglas B. Currivan of RTI International.

www.chis.ucla.edu

This report describes the weighting and variance estimation methods used in CHIS 2015-2016. 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 2015-2016 Methodology Series: Report 5 - Weighting and Variance Estimation*. Los Angeles, CA: UCLA Center for Health Policy Research, 2017.

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 2015-2016 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, San Diego County Health and Human Services Agency, the Williams Institute at UCLA Law School, Marin County Health and Human Services, Imperial County Public Health Department, and the California Immigration Research Initiative.

PREFACE

Weighting and Variance Estimation is the fifth and final in a series of methodological reports describing the 2015-2016 California Health Interview Survey (CHIS). 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. RTI International was responsible for data collection and the preparation of five methodological reports from the 2015-2016 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 Report Series for CHIS 2015-2016

The methodological reports for CHIS 2015-2016 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 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 2015-2016. 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.

For further methodological details not covered in this report, refer to the other methodological reports in the series at http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx. General information on CHIS data can be found on the California Health Interview Survey Web site at http://www.chis.ucla.edu or by contacting CHIS at CHIS@ucla.edu.

TABLE OF CONTENTS

List of Tables

1. CHIS 2015-2016 SAMPLE DESIGN AND METHODOLOGY SUMMARY

1.1 Overview

A series of five methodology reports are available with more detail about the methods used in CHIS 2015-2016.

- 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. For methodology reports from previous CHIS cycles, go to http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx

The CHIS is a population-based telephone survey of California's residential, non-institutionalized population conducted every other year since 2001 and continually beginning in 2011. CHIS is the nation's largest state-level health survey and one of the largest health surveys in the nation. The UCLA Center for Health Policy Research (UCLA-CHPR) conducts CHIS 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 and optimized to meet 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 racial and 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 (visit UCLA-CHPR's publication page at http://healthpolicy.ucla.edu/publications/Pages/default.aspx for examples of CHIS studies).

1.2 Switch to a Continuous Survey

From the first CHIS cycle in 2001 through 2009, CHIS data were collected during a 7- to 9-month period every other year. Beginning in 2011, CHIS data have been collected continually over a 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 2015 data were collected between May 2015 and mid-February 2016. CHIS 2016 data were collected between January and December 2016. Approximately half of the interviews were conducted during the 2015 calendar year and half during the 2016 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 year for the age group corresponding to the data file in use (adult, adolescent, or child). In CHIS 2015-2016, data users will be able to produce single-year estimates using the weights provided (referred to as CHIS 2015 and CHIS 2016, respectively).

See what's new in the 2015-2016 CHIS sampling and data collection here:
http://healthpolicy.ucla.edu/chis/design/Documents/whats-new-chis-2015-2016.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 statistical 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 the *CHIS 2015-2016 Methodology Series: Report 5—Weighting and Variance Estimation* posted 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.

1.3 Sample Design Objectives

The CHIS 2015-2016 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 racial and 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 RDD sample was designed to achieve the required number of completed adult interviews by using approximately 50% landline and 50% cellular phone numbers. For the 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 same geographic stratification of the state has been used since CHIS 2005. The Los Angeles County stratum included eight sub-strata for Service Planning Areas, and the San Diego County stratum included six sub-strata for Health Service Districts. Most of the strata (39 of 44) consisted of a single county with no sub-strata (see counties 3-41 in Table 1-1). Three multi-county strata comprised the 17 remaining counties (see counties 42-44 in Table 1-1). A sufficient number of adult interviews were allocated to each stratum and sub-stratum to support the first sample design objective for the two-year period—to provide health estimates for adults at the local level. Asian surname sample list frames added 426 Japanese, 280 Korean, and 359 Vietnamese adult interviews based on self-identified ethnicity for the combined 2015 and 2016 survey years.[1] Additional samples from both the landline and cell phone frames produced 1,042 interviews in 2015 within Marin County and 2,388 interviews in 2016 within San Diego County. Furthermore, an address-based sample from the USPS Delivery Sequence File produced 258 landline or cell phone interviews in 2016 within the northern part of Imperial County.

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 of the randomly

[1] For the 2015 and 2016 survey years combined, all sample frames produced totals of 667 Japanese, 497 Korean, and 597 Vietnamese adult interviews.

selected parent/guardian were randomly selected; the adolescent was interviewed directly, and the adult sufficiently knowledgeable about the child's health completed the child interview.

The CHIS RDD 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). However, given the smaller sample sizes of one-year data files, two or more pooled cycles of CHIS data are generally required to produce statistically stable estimates for small population groups such as racial/ethnic subgroups, children, teens, etc. 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 increase the sample size further for Koreans and Vietnamese. Surname and given name lists were used similarly to increase the yield of Californians of Japanese descent.

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 2015 and 2016, the goal was to complete approximately 50% of all RDD interviews statewide with adults contacted via cell phone. Because the geographic information available for cell phone numbers is limited and not as precise as that for landlines, cell phone numbers were assigned to the same 44 geographic strata (i.e., 41 strata defined by a single county and 3 strata created by multiple counties) using a classification associated with the rate center linked to the account activation. 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. The adult owner of the sampled cell phone number was automatically selected for CHIS. Cell numbers used exclusively by children under 18 were considered ineligible. A total of 1,594 teen interviews and 4,293 child interviews were completed in CHIS 2015-2016 with approximately 58% coming from the cell phone sample.

The cell phone sampling method used in CHIS has evolved significantly since its first implementation in 2007 when only cell numbers belonging to adults in cell-only households were eligible for sampling adults. These changes reflect the rapidly changing nature of cell phone ownership and use in the US.[2] There have been three 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

[2] http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201605.pdf

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 third, enacted in CHIS 2015-2016 was to increase the fraction of the sample comprised of cell phones from 20% to 50% of completed interviews.

Table 1-1. California county and county group strata used in the CHIS 2015-2016 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,
4. Santa Clara	24. Placer	Lassen, Modoc, Trinity, Del Norte
5. San Bernardino	25. Merced	44. Mariposa, Mono, Tuolumne,
6. Riverside	26. Butte	Alpine, Amador, Calaveras, Inyo

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

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 seven geographic sampling strata created using telephone area codes. In CHIS 2009 and 2011-2012, the number of cell phone strata was increased to 28. These strata were created using both area codes and the geographic information assigned to the number. Beginning in CHIS 2011, with the availability of more detailed geographic information, the number of strata was increased to 44 geographic areas that correspond to single and grouped counties similar to the landline strata. The use of 44 geographic strata continued in CHIS 2015-2016.

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 Tagalog. Tagalog interviews were conducted for part of the CHIS 2013-2014 cycle, but 2015-2016 were the first cycle years that Tagalog interviewers were conducted from the beginning of data collection. 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.

RTI International designed the methodology and collected data for CHIS 2015-2016, under contract with the UCLA Center for Health Policy Research. RTI is an independent, nonprofit institute that provides research, development, and technical services to government and commercial clients worldwide, with specialization in designing and implementing large-scale sample surveys. For all sampled households, RTI 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. If 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 2015-2016 by the type of sample (landline RDD, surname list, cell RDD, and ABS). 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.

Interviews in all languages were administered using RTI's computer-assisted telephone interviewing (CATI) system. The average adult interview took about 41 minutes to complete. The average child and adolescent interviews took about 19 minutes and 22 minutes, respectively. For "child-first" interviews, additional household information asked as part of the child interview averaged about 12 minutes. Interviews in non-English languages typically took somewhat longer to complete. More than 13 percent of the adult interviews were completed in a language other than English, as were about 24 percent of all child (parent proxy) interviews and 25 percent of all adolescent interviews.

Table 1-2. Number of completed CHIS 2015-2016 interviews by type of sample and instrument

Type of sample[1]	Adult[2]	Child	Adolescent
Total all samples	42,089	4,293	1,594
Landline RDD	15,106	1,178	542
Vietnamese surname list	3,558	316	111
Korean surname list	1,772	130	64
Japanese surname list	631	34	25
Cell RDD	19,722	2,521	807
Marin County Oversample[3]	1,042	83	33
Imperial County ABS Oversample	258	31	12

[1] Completed interviews listed for each sample type refer to the sampling frame from which the phone number was drawn. Interviews could be conducted using numbers sampled from a frame with individuals who did not meet the target criteria for the frame but were otherwise eligible residents of California. Interviews from the Marin County oversample include respondents who did not live in this county and interviews from the Vietnamese, Korean, or Japanese surname lists include respondents who do not have one of these ethnicities. For example, only 182 of the 3,558 adult interviews completed from the Vietnamese surname list involved respondents who indicated being having Vietnamese ethnicity.

[2] Includes interviews meeting the criteria as partially complete,

[3] Completed interviews for the Marin County oversample do not include interviews completed via the Vietnamese surname list frame. These interviews are counted in the row for the Vietnamese surname list.

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

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

Table 1-3. CHIS 2015-2016 survey topic areas by instrument

Health status	**Adult**	**Teen**	**Child**
General health status	✓	✓	✓
Days missed from school due to health problems		✓	✓
Health-related quality of life (HRQOL)	✓	✓	
Health conditions	**Adult**	**Teen**	**Child**
Asthma	✓	✓	✓
Diabetes, gestational diabetes, pre- /borderline diabetes	✓		
Heart disease, high blood pressure, stroke	✓		
Physical, behavioral, and/or mental conditions			✓
Physical disabilities, blindness, deafness	✓		
Mental health	**Adult**	**Teen**	**Child**
Mental health status	✓	✓	
Perceived need, access and utilization of mental health services	✓	✓	
Suicide ideation and attempts	✓	✓	
Functional impairment, stigma	✓		
Health behaviors	**Adult**	**Teen**	**Child**
Dietary intake, fast food and soda intake	✓	✓	✓
Water Consumption		✓	
Physical activity and exercise, commute from school to home		✓	✓
Sedentary time		✓	✓
Walking for transportation and leisure	✓		
Doctor discussed nutrition/physical activity		✓	✓
Flu Shot	✓	✓	✓
Alcohol use	✓	✓	
Cigarette and E-cigarette use	✓	✓	
Sexual behavior	✓	✓	
Breastfeeding			✓
Women's health	**Adult**	**Teen**	**Child**
Mammography screening	✓		
Pregnancy	✓		
Dental health	**Adult**	**Teen**	**Child**
Last dental visit, main reason haven't visited dentist	✓	✓	✓

(continued)

Table 1-3. CHIS 2015-2016 survey topic areas by instrument (continued)

Neighborhood and housing	**Adult**	**Teen**	**Child**
Safety, social cohesion	✓	✓	✓
Homeownership, length of time at current residence	✓		
Park use		✓	✓
Civic engagement	✓	✓	
Building Healthy Communities	✓		
Access to and use of health care	**Adult**	**Teen**	**Child**
Usual source of care, visits to medical doctor	✓	✓	✓
Emergency room visits	✓	✓	✓
Delays in getting care (prescriptions and medical care)	✓	✓	✓
Medical home, timely appointments, hospitalizations	✓	✓	✓
Developmental screening			✓
Communication problems with doctor	✓		✓
Internet use for health information	✓		✓
Tele-medical care	✓		
Family planning	✓		
Change of usual source of care	✓		
Food environment	**Adult**	**Teen**	**Child**
Access to fresh and affordable foods	✓		
Where teen/child eats breakfast/lunch, fast food at school		✓	✓
Availability of food in household over past 12 months	✓		
Hunger	✓		
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	✓	✓	✓
Difficulty finding private health insurance	✓		
High deductible health plans	✓	✓	✓
Partial scope Medi-Cal	✓		

(continued)

Table 1-3. CHIS 2015-2016 survey topic areas by instrument (continued)

Public program eligibility	**Adult**	**Teen**	**Child**
Household poverty level	✓		
Program participation (CalWORKs, Food Stamps, SSI, SSDI, WIC, TANF)	✓	✓	✓
Assets, alimony/child support, social security/pension, worker's compensation	✓		
Medi-Cal and Healthy Families eligibility	✓	✓	✓
Reason for Medi-Cal non-participation among potential beneficiaries	✓	✓	✓
Bullying and interpersonal violence	**Adult**	**Teen**	**Child**
Bullying, personal safety, school safety, interpersonal violence		✓	
Parental involvement/adult supervision	**Adult**	**Teen**	**Child**
Adult presence after school, role models, resiliency		✓	
Parental involvement		✓	
Child care and school attendance	**Adult**	**Teen**	**Child**
Current child care arrangements			✓
Paid child care	✓		
Preschool/school attendance, name of school		✓	✓
Preschool quality			✓
School instability		✓	
First 5 California: "Talk, Read, Sing Program"			✓
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	✓		
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	✓		
Education, English language proficiency	✓		
Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home	✓	✓	✓
Education of primary caretaker			✓
Citizenship, immigration status, country of birth, and length of time in U.S. of parents			✓

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

1.5 Responsive and Adaptive Design Elements

The CHIS 2015 and 2016 data collection protocol included the following two responsive design protocols to maximize response rates, provide protection against nonresponse bias, and control data collection costs:

1) a propensity model experiment in the first phase of each quarterly data collection that identified a set of cases with low propensities to discontinue calling for the remainder of Phase 1
2) a second nonresponse follow-up (NRFU) phase in each quarterly data collection period where a different protocol was implemented to increase response rates and reduce the risk of nonresponse bias.

Additional documentation on the responsive design protocols and outcomes is available in the *CHIS 2015-2016 Methodology Series: Report 2—Data Collection Methods* posted at http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx.

1.6 Response Rates

The overall response rates for CHIS 2015 and 2016 are composites 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). For CHIS 2015, the landline/list sample household response rate was 9.1 percent (the product of the screener response rate of 21.0 and the extended interview response rate at the household level of 43.2 percent). The cell sample household response rate was 9.8 percent, incorporating a screener response rate of 21.5 percent household-level extended interview response rate of 45.9 percent. For CHIS 2016, the landline/list sample household response rate was 6.8 percent (the product of the screener response rate of 15.5 and the extended interview response rate at the household level of 44.0 percent). The cell sample household response rate was 8.4 percent, incorporating a screener response rate of 18.5 percent household-level extended interview response rate of 45.4 percent. CHIS uses AAPOR response rate RR4 (see more detailed in *CHIS 2015-2016 Methodology Series: Report 4 – Response Rates*).

Within the landline and cell phone sampling frames for 2015, the extended interview response rate for the landline/list sample varied across the adult (41.8 percent), child (44.7 percent) and adolescent (17.1 percent) interviews. For 2016, the extended interview response rate for the landline/list sample varied across the adult (41.3 percent), child (69.6 percent) and adolescent (17.9 percent) interviews. The

adolescent rate includes the process of obtaining permission from a parent or guardian. The adult interview response rate for the cell sample was 48.5 percent, the child rate was 43.9 percent, and the adolescent rate was 17.4 percent in 2015 (see Table 1-4a). The adult interview response rate for the cell sample was 46.9 percent, the child rate was 59.7 percent, and the adolescent rate was 21.6 percent in 2016 (see Table 1-4c). Multiplying these rates by the screener response rates used in the household rates above gives an overall response rate for each type of interview for each survey year (see Table 1-4b and Table 1-4d, respectively). As in previous years, household and person level response rates vary by sampling stratum. CHIS response rates are similar to, and sometimes higher than, other comparable surveys that interview by telephone.

Table 1-4a. CHIS 2015 response rates – Conditional

Type of sample	Screener	Household	Adult (given screened)	Child (given screened)	Adolescent (given screened & permission)
Overall	21.4%	45.2%	47.2%	44.0%	17.3%
Landline RDD	21.0%	43.2%	41.8%	44.8%	17.1%
Cell RDD	21.5%	45.9%	48.5%	43.9%	17.4%

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Table 1-4b. CHIS 2015 response rates – Unconditional

Type of sample	Screener	Household	Adult (given screened)	Child (given screened)	Adolescent (given screened & permission)
Overall	21.4%	9.7%	10.1%	9.4%	3.7%
Landline RDD	21.0%	9.1%	8.8%	9.4%	3.6%
Cell RDD	21.5%	9.8%	10.4%	9.4%	3.7%

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Table 1-4c. CHIS 2016 response rates – Conditional

Type of sample	Screener	Household	Adult (given screened)	Child (given screened)	Adolescent (given screened & permission)
Overall	17.8%	45.1%	44.6%	63.0%	20.0%
Landline RDD	15.5%	44.0%	41.3%	69.6%	17.9%
Cell RDD	18.5%	45.4%	46.9%	59.7%	21.6%

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Table 1-4d. CHIS 2016 response rates – Unconditional

Type of sample	Screener	Household	Adult (given screened)	Child (given screened)	Adolescent (given screened & permission)
Overall	17.8%	8.0%	7.9%	11.2%	3.6%
Landline RDD	15.5%	6.8%	6.4%	10.8%	2.8%
Cell RDD	18.5%	8.4%	8.7%	11.1%	4.0%

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

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 34.5 percent of the landline RDD sample telephone numbers not identified by the sample vendor as business numbers or not identified by RTI's dialer software as nonworking numbers, and for 92.3 percent of surname list sample numbers. Combining these two frames, advance letters were sent to 40.5 percent of all fielded landline telephone 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 2015-2016 advance letter to encourage cooperation. Additional incentives were offered to cell phone and Phase 2 nonresponse follow up (NRFU) respondents. Details on the incentives are provided in Table 1-5.

Table 1-5. 2015-2016 CHIS incentives by interview type

Type of interview	Adult
Cell Phone Screener	$5
Cell Phone Adult Interview	$20
Cell Phone Child Interview	$10
Cell Phone Teen Interview	$10
Nonresponse Follow-Up Adult Interview	$40
Nonresponse Follow-Up Child Interview	$20
Nonresponse Follow-Up Teen Interview	$20

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

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.

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. In the 2015-2016 CHIS, either a spouse/partner or adult child completed a proxy interview for 274 adults. A reduced questionnaire, with questions identified as appropriate for a proxy respondent, was administered.

Further information about CHIS data quality and nonresponse bias is available at http://healthpolicy.ucla.edu/chis/design/Pages/data-quality.aspx.

1.7 Weighting the Sample

To produce population estimates from CHIS data, weights were 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 was weighted to represent the non-institutionalized population for each sampling stratum and statewide. The weighting procedures used for CHIS 2015-2016 accomplish the following objectives:

- Compensate for differential probabilities of selection for phone numbers (households) and persons within household;
- Reduce biases occurring because nonrespondents may have different characteristics than respondents;
- Adjust, to the extent possible, for undercoverage in the sampling frames and in the conduct of the survey;
- Reduce the variance of the estimates by using auxiliary information; and
- Account for the second-phase sampling that was part of the responsive and adaptive design (Phase 2 NRFU).

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 was used to compute a person-level weight, which includes adjustments for the within-household sampling of persons and for nonresponse. The final step was to adjust the person-level weight using weight calibration, a procedure that forced the CHIS weights to sum to estimated population control totals simultaneously from an independent data source (see below).

Population control totals of the number of persons by age, race, and sex at the stratum level for CHIS 2015-2016 were created primarily from the California Department of Finance's (DOF) 2015 and 2016 Population Estimates, and associated population projections. The procedure used several 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), and education. 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 calibration.

The DOF control totals used to create the CHIS 2015 and 2016 weights are based on 2010 Census counts, as were those used for the 2013-2014 cycle. Please pay close attention when comparing estimates using CHIS 2015-2016 data with estimates using data from CHIS cycles before 2010. The most accurate California population figures are available when the U.S. Census Bureau conducts the decennial census. For periods between each census, population-based surveys like CHIS must use population projections based on the decennial count. 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.

1.8 Imputation Methods

Missing values in the CHIS data files were replaced through imputation for nearly every variable. This was a substantial task designed to enhance the analytic utility of the files. RTI 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 RTI 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. The hot deck approach is one of the most commonly used methods for assigning values for missing responses. Using a hot deck, a value reported by a respondent for a specific item was 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 answered a survey item formed a pool of donors, while the item nonrespondents formed a group of recipients. A recipient was matched to the subset pool of donors based on household and individual characteristics. A value for the recipient was then randomly imputed from one of the donors in the pool. RTI used hot deck imputation to impute the same items that have been imputed 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 RTI and some sensitive variables for which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2015 and CHIS 2016 were low, with most variables missing valid responses for less than 1% of the sample.

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 without donor replacement was used. This method replaced 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 corresponded to the nature of the variable being imputed (e.g. linear regression for continues variables, logistic regression for binary variables, etc.). Donors and recipients were 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 RTI. UCLA-CHPR began their imputation process by imputing household income and educational attainment, so that these characteristics are available for the imputation of other variables. Sometimes 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.

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

2. WEIGHTING ADJUSTMENTS

Researchers apply analysis weights to survey responses to produce estimates for the target population. The weights are designed to produce estimates with minimal biases and maximal precision (i.e., relatively small standard errors). This section provides an overview of the weighting methodology used for the CHIS 2015-2016 one-year and two-year weights.

Specifically, the approach to weighting CHIS data is provided in Section 2.1. Base weights and adjustments are combined to form the CHIS analysis weights. The weight components are listed in Section 2.2, along with a link to the section of this report where details are provided. Differences in the CHIS 2015-2016 nonresponse adjustments from prior years are also discussed. Because CHIS includes multiple sampling frames, Section 2.3 contains an overview of procedures to blend multiple samples within a single year of the study. Similar procedures were used to produce analysis weights for the two-year CHIS 2015-2016 data file (Section 2.4). This chapter concludes in Section 2.5 with a brief discussion of quality assurance procedures.

2.1 Weighting Approach

The weighting approach used for CHIS 2015-2016 follows the paradigm set in prior rounds of the study. Specifically, the methods to construct the weights follow standard design-based techniques that account for sampling from multiple frames. The use of multiple frames—landline, cell, and surname—has been used consistently since CHIS 2009 to ensure coverage of the residential California population. Additionally, as with CHIS 2013-2014, an address-based sample (ABS) was selected to supplement the sample of telephone numbers. In CHIS 2016, an ABS sample was chosen for a targeted area within Imperial County.

Not only do CHIS weights account for differential sampling by frame, but they also include adjustments to combine across these frames. These procedures resulted in a set of unified analysis weights applicable for all analyses. For example, these weights are used to generate estimates at the state-level as well as sub-state estimates at the county level.

One set of weights was produced for each CHIS person-level interview: adult, child and teen. Each weight was constructed to address the following nuances of the design and data collection actualities attributed to each interview:

- Differential selection probabilities of sampled households by telephone and address frame across design strata, for households included in a nonresponse follow-up, and for persons within the selected households;
- Reduce bias that may occur in the estimates when nonrespondents differ from their respondent counterparts;
- Reduce coverage bias associated with differences of the respondent distributions from the intended target population; and
- Improve the precision of CHIS estimates (i.e., small standard errors) by adjusting to population information and adjusting any outlier weights.

An overview of the specific weight components is provided in Section 2.2

As discussed in Chapter 9, estimates for the target population are produced only if analyses account for the CHIS sampling design and the weights. Ignoring either the sampling design or the analysis weights is not recommended.

2.2 Weighting Adjustments

CHIS one-year analysis weights were developed for adult, child and teen completed interviews. The weights were constructed as a function of an initial base weight (inverse selection probability within sampling frame and design stratum) multiplied by a sequential series of adjustments to address nonresponse, subsampling, unknown eligibility, and differential coverage from the intended target population. The adjustments are summarized in Section 2.2.1, followed by a comparison of nonresponse adjustment methods for CHIS 2015-2016 and prior years (Section 2.2.2).

2.2.1 Components of the CHIS Analysis Weights

Details of the one-year weight components are provided in Chapters 3-6, beginning with the household weight (Chapter 3).

The weight associated with the selected household was derived as the product of the following components:

- base weights defined by data collection period (quarter), sampling frame and design stratum (Section 3.1)
- adjustment for sampling associated with the Korean, Vietnamese, and Japanese surname list frames (Section 3.2)

- adjustment to combine samples across quarter and sampling frame (Section 3.3)
- adjustment for households subsampled for a nonresponse follow-up (Section 3.4)
- adjustment for household without a known study-eligibility status (Section 3.5)
- adjustment for nonresponse to the CHIS household screener (Section 3.6)

The final household weight was used as the basis for three analysis weights (adult, child and teen) corresponding to extended interviews. The adult analysis weights (Chapter 4) was constructed as the final household weight multiplied by the following adjustments:

- inverse selection probability of one adult within each household with a completed screener (Section 4.1)
- adjustment for nonresponse follow-up subsampling where the sampled adult participated in the initial recruitment phase (Section 4.2)
- adjustment for adult nonresponse (Section 4.3)
- adjustment to align the weight sums to population counts by telephone-usage status (Section 4.4)
- adjustment to combine cell phone and landline samples (Section 4.5)
- adjustment to combine the cell/landline sample with a supplemental sample in Imperial County (Section 4.6)
- adjustment to align the weight sums to adult population counts by geographic area within California, demographic characteristics, and other such information (Section 4.7)

Note that samples were selected from design strata but final weight adjustments were applied within the reported stratum from the adult interview. Differences between design and reported strata were most apparent for the cell phone sample because these numbers were more likely retained when persons relocated to a different county. Where applicable, tables in this report will include information to identify which stratum type was used in the analysis.

Like the adult weights, the child analysis weights (Chapter 5) was constructed as the final household weight multiplied by the following adjustments:

- adjustment for the inability to sample one child per household because of adult nonresponse (Section 5.1)
- inverse selection probability of one child within each household with a completed screener (Section 5.2)

- adjustment for child-interview nonresponse associated with a knowledgeable parent (Section 5.3)
- adjustment to align the weight sums to population counts by telephone-usage status (Section 5.4)
- adjustment to combine cell phone and landline households with children (Section 5.5)
- adjustment to combine the cell/landline sample with a supplemental sample in Imperial County (Section 5.6)
- adjustment to align the weight sums to child population counts by geographic area within California, demographic characteristics, and other such information (Section 5.7)

The teen analysis weights (Chapter 6) was constructed in a similar fashion as the product of the final household weight and the following adjustments:

- inverse selection probability of one teen within each household with a completed screener (Section 6.1)
- adjustment for nonresponse linked to the parental permission or to the teen (Section 6.2)
- adjustment to align the weight sums to population counts by telephone-usage status (Section 6.3)
- adjustment to combine cell phone and landline households with one or more teens (Section 6.4)
- adjustment to combine the cell/landline sample with a supplemental sample in Imperial County (Section 6.5)
- adjustment to align the weight sums to teen population counts by geographic area within California, demographic characteristics, and other such information (Section 6.6)

The second to last section of each Chapters 4-6 describes the final analysis weight which sometimes included constraints on outlier weights (Section 2.5). The final chapter for the person-level weights summarizes an adjustment to combine across the one-year files for CHIS 2015-2016.

A calibration adjustment (Kott, 2006; Valliant et al., 2013), such as those discussed for the adult weights in Sections 4.4 and 4.7, was applied to align the CHIS weights to population counts, also referred to as calibration controls or control totals. Because control totals for the CHIS target population by key covariates (e.g., design stratum) did not exist, the population counts needed to be estimated from existing information. The procedures to calculate the estimated control totals followed those used in prior rounds of CHIS and are detailed in Chapter 7.

Analysis weights address bias associated with unit nonresponse that occurs when a sample member either declines to participate or when they do not provide sufficient information for analyses. A CHIS sample member needed to complete the interview at least through the end of Section K to be classified as a respondent. Some respondents, however, declined to provide information to critical items needed for the creation of the analysis weights. This missing information was supplied through various imputation procedures detailed in Chapter 8 after the data were processed (see *CHIS 2015-2016 Methodology Series: Report 3 - Data Processing Procedures*).

Chapter 9 contains a discussion on variance estimation for CHIS 2015-2016. This includes Taylor Series linearization calculated with a single set of analysis, and balanced repeated replicate variance estimation calculated with a series of (replicate) weights. Software to calculate estimated standard errors are also discussed.

This report contains two supplementary appendices. Appendix A consists of a series of tables with frame counts, sample sizes, and base weights by the design strata. Appendix B provides summary statistics for each component discussed above.

2.2.2 Raking vs. Model-based adjustments for Nonresponse

In past CHIS cycles, a weighting class adjustment, much like those discussed previously, was used to account for screener and extended-interview nonresponse. Weighting classes (i.e., groups) were formed by combining binary, categorical, or categorized continuous variables thought to be associated with response and preferably also with characteristics of importance from the study. As noted in Kim et al., (2007), use of many variables can result in too many or even small (empty) weighting classes that hinder the calculation of an efficient nonresponse-adjusted weight. Determining an effective mechanism for collapsing small cells can be a time-consuming process, yielding minimal gains in precision (via reduced variations in weights) and possibly limiting the reduction of bias attributable to nonresponse. Consequently, incorporating only a few variables limits the capacity to reduce nonresponse bias, the true goal of this weight adjustment. Therefore, in CHIS 2015-2016, a model-based approach was implemented with the SUDAAN® WTADJUST procedure (RTI, 2012).

PROC WTADJUST enables the creation of a model-based adjustment that can be used either directly as implemented in CHIS 2015-2016 or to form weighting classes by categorizing the response propensities into at least five groups. Additionally, for the direct method, the procedure enables adjustments to be constrained within a defined set of values to maintain control over the variation of the

weight adjustments or even the subsequent adjusted weights. We built the associated models based on an examination of the response patterns by interview type (LS, CP, or ABS).

Candidate nonresponse model variables were identified from among those available for respondents and nonrespondents at each adjustment step (e.g., adult respondents and sampled adults with only a completed screener). Only variables with a significant test of association with the response indicator, used in a similar nonresponse adjustment in prior rounds, or associated with the sampling design were retained for further evaluation. Once compiled, the set of covariate were evaluated simultaneously via a classification and regression tree (CART) analyses (Breiman et al., 1984). The goal of CART software in this case is to identify a parsimonious model that predicts response. Initially, the input variable with the most predictive power is retained; interactions with of the other covariates are evaluated to determine the next set of significant "tree leaves". This iterative process continues until stopping rules are met (e.g., group sizes must meet or exceed 50 sample members). The resulting set of model covariates may include the interactions of a subset of variable values with others instead of the interaction of all levels as with raking.

2.3 Combining CHIS Samples within a Year

Multiple CHIS samples instead of one were drawn throughout a calendar year to:

- maximize coverage of the target population by selecting landline and cellular telephone numbers;
- provide current contact information for active telephone numbers;
- allow differential sampling by design stratum to address updates to projected response rates; and
- incorporate real-time requests for supplemental samples.

Therefore, procedures were required to combine the samples produce a single set of weights for estimation of population quantities. Methods to address samples from landline and cellular telephone sampling frames are discussed in Section 2.3.1. A similar methodology to incorporate a sample of addresses is discussed in Section 2.3.2.

2.3.1 Combining Cellular and Landline Telephone Samples

For several years, the CHIS sampling design has required samples drawn from both landline and cellular list frames. However, both types of telephone numbers may reach a proportion of same

households — 50.7% of California households were estimated in 2015 to have a landline telephone number in addition to at least one cell phone (Blumberg and Luke, 2016). Conversely, only 5.5% of California households in 2015 were estimated to be landline only (Blumberg and Luke, 2016). Hence, the wise decision to convert CHIS from a landline-only design to a dual-frame design with sample of both landline and cell numbers. Otherwise, estimates from CHIS under a landline-only design would underestimate the proportion of cell-only households, and vice versa.

The benefits of a dual-frame design in increasing coverage of the target population induces complexity for estimation. Estimates from the landline sampling frame include portions associated with landline-only households plus those with both landline and cell phone access (dual users). Similarly, estimates from the cellular sampling frame include portions associated with cell-only households plus dual users. Combining the two estimates would conceptually over-represent estimates from dual-use households by as much as two times. Therefore, weighting adjustments are needed to align the two samples to collectively represent the target population.

The CHIS weighting methodology follows work of Hartley (1962) to combine estimates from the same conceptual population. Let Y represent the characteristic of interest for a research project. Define θ_L and θ_C to be the population parameters for the landline and cellular household population in California for a statistic θ (e.g., total, mean). A CHIS estimate from the landline sampling frame would produce $\hat{\theta}_L = \hat{\theta}_{L\bullet} + \hat{\theta}_{LC}^{(L)}$, where $\hat{\theta}_{L\bullet}$ is the estimate from landline-only households and $\hat{\theta}_{LC}^{(L)}$ is the frame-specific estimate for the dual users. The associated unadjusted estimate from the cellular frame is $\hat{\theta}_C = \hat{\theta}_{\bullet C} + \hat{\theta}_{LC}^{(C)}$, where $\hat{\theta}_{\bullet C}$ is the estimate from cell-only households and $\hat{\theta}_{LC}^{(C)}$ is the associated estimate for dual-users. The combined estimate, accounting for the overlap in the frames, should be estimated as

$$\begin{aligned}\hat{\theta} &= \hat{\theta}_{L\bullet} + \lambda\hat{\theta}_{LC}^{(L)} + \hat{\theta}_{\bullet C} + (1-\lambda)\hat{\theta}_{LC}^{(C)} \\ &= \hat{\theta}_{L\bullet} + \hat{\theta}_{LC} + \hat{\theta}_{\bullet C}\end{aligned} \tag{2.1}$$

where λ is the composite factor less than one such that the combined dual-use estimates from both frames is linked to the estimated number of households accessed by a landline and cellular telephone $\hat{\theta}_{LC}$. With this approach, the composite factor (λ) for CHIS was applied to the weights to enable analyses and allowed to differ by design stratum. Details for the adult adjustment to combine landline and cell phone samples is found in Section 4.5, the child adjustment in Section 5.5, and the teen adjustment in Section 6.4.

2.3.2 Combining Telephone and Address Samples

For CHIS 2016, an addressed-based sample (ABS) was chosen for a supplemental sample within a targeted area of Imperial County (see *CHIS 2015-2016 Methodology Series: Report 1 – Sample Design*). A single set of weights (linear and replicate) was created to combine Imperial County respondents obtained from the ABS sample and the telephone sample (landline, cell, and surname). The methodology was the same as discussed in the previous section. Details for the adult adjustment is found in Section 4.6, the child adjustment in Section 5.6, and the teen adjustment in Section 6.5.

2.3.3 Combining CHIS Samples across Years

Analysis weight for analyzing the CHIS 2015-2016 two-year file were constructed with the same methodology as described in Section 2.3. Specifically, a weight adjustment was applied to capture the composite factor for the estimation equation,

$$\hat{\theta} = \lambda \hat{\theta}^{(2015)} + \left(1 - \lambda\right) \hat{\theta}^{(2016)} \tag{2.2}$$

the sum of the one-year estimates and the composite factor. Details for the adult adjustment is found in Section 4.9, the child adjustment in Section 5.9, and the teen adjustment in Section 6.8.

2.4 Quality Checks

A series of quality control procedures was implemented at each step to ensure the accuracy of survey weights. A few examples are provided below.

First, the weight sums were compared before and after an adjustment, and after all the weighting steps, against external counts such as those tabulated from the American Community Survey. Large differences would have indicated either errors or potential problems in model-based adjustments.

Statistics of the weights (e.g., variance, minimum, maximum) were compared before and after an adjustment. Large differences have signaled a need for further review. For example, a large relative change in an unequal weighting effect (UWE; i.e., design effect associated with the weights) calculated by important domains (e.g., race/ethnicity or geographic location) would be evaluated to determine if additional variables should be used for the weight-adjustment model or if WTADJUST bounds on the adjustments should be tightened.

The weights were also examined for outliers using the 3×interquartile range rule, overall and within key domains (see, e.g., Chen et al., 2014). Outliers were subject to trimming only after a thorough review of the weight components.

At each stage of the weighting process, sums of the replicate weights (Section 9) were compared against the corresponding value for the linear weights; this step ensured that approximately half of the replicate values were at or below the linear value. Estimated standard errors using linear and replicate weights were evaluated where large differences would require further evaluation of both sets of weights.

3. HOUSEHOLD WEIGHTING

The first stage of selection for CHIS 2015-2016 as in prior years was the household by way of a sampled landline telephone number, a sampled cell phone number, or a sampled address specifically for a supplemental Imperial County address-based sample (ABS). Additional details on the CHIS sample design is available in *CHIS 2015-2016 Methodology Series: Report 1—Sample Design*.

Weights generated at this stage in the process are called "household weights" to keep with the historic CHIS label. These weights by themselves, however, should not be used to generate estimates for the household population in California. Primarily, they do not incorporate important adjustment factors related to nonresponse within the household nor calibration to the number of households by county.

In this chapter, we detail the steps used to calculate the household-level weight by type of sampling frame—landline (RDD and surname), cell, and ABS. Differences by year within CHIS 2015-2016 are noted where appropriate. We use the finalized weight as the basis for the person-level analysis weights—adult, child (proxy), and teen—discussed in the subsequent chapter of this report.

Specifically, we define the initial base weights by sampling frame in Section 3.1 that accounts for the first of two phases of sampling at the household level. Section 3.2 contains an adjustment specific to surname sampling within CHIS 2015. A unifying adjustment is discussed in Section 3.3 to combine multiple samples within each year of the study. We define a subsampling adjustment for the second phase of the design in Section 3.4. As discussed in Section 3.5, weights were adjusted for unknown residential status and then, among residences, for those without a known study eligibility status. Weights for those with unknown residential status or unknown eligibility status were then set to zero. Next, we applied an adjustment for household-level nonresponse defined as households without a completed screener (Section 3.6). The final household weight is defined in Section 3.7.

Frame size, sample size and base weight by sampling frame and design stratum are provided in Appendix A. Statistics for the adjustments and the final weight are provided in Table B-1 in Appendix B.

3.1 Base Weights

A base weight, also referred to as a "design weight" or "sampling weight", adjusts only for the specific process of sampling from each sampling frame. A phone number was our proxy for household for those other than the ABS sample. The base weight was calculated as the inverse of the selection probability for each sampled record (phone number or address) from the respective frame of all phone numbers or addresses and by data collection quarter.

For clarity and convenience of notation, we provide the following indicators for the various sampling frames:

- LL = random-digit-dial (RDD) landline telephone sampling frame
- SV = Vietnamese surname (landline) telephone sampling frame
- SK = Korean surname (landline) telephone sampling frame
- SJ = Japanese surname (landline) telephone sampling frame
- CE = RDD cellular telephone sampling frame
- ABS = addressed-based sampling frame

The base weights were calculated as follows in keeping with the stratified simple random sampling design within sampling frame:

$$HW0_{FQhi} = \frac{N_{FQh}}{n_{FQh}} \tag{3.1}$$

where F indexes the sampling frame; Q indexes the quarter of data collection; h indexes the design stratum; i indexes the household (telephone or address); n_{FQh} is the size of the sample selected within stratum h from sampling frame F within quarter Q; and N_{FQh} is the associated frame count. Note that the household index is suppressed from the sample and frame size notation above (n_{FQh} and N_{FQh}, respectively) because the values were identical across households within the same design stratum.

For CHIS 2015, two sample vendors were used to evaluate possible differences in their frame methodologies. Only one vendor provided samples for CHIS 2016. For simplicity, we suppress a subscript to indicate base weights by vendor that is appropriate only to one year within this sequence.

3.2 Korean, Vietnamese, and Japanese Surname List Adjustment

Surname sampling was introduced for CHIS 2015 after creation of the landline sampling frame and after the start of data collection. Therefore, the surname list frames were a proper subset of the RDD landline frame. Thus, the surname list sample could have been chosen either from the respective surname frame or from LL.

In contrast, landline sampling frames were generated for CHIS 2016 that excluded the surname landline telephone numbers. Consequently, landline and surname samples were drawn from mutually exclusive sampling frames. This design change was made in attempt to enable comparisons among respondents sampled from the respective frames.

To address the fact that surname samples had twice the chance of being selected for CHIS 2015 but not for CHIS 2016, the following adjusted base weight was created:

$$HW1_{FQhi} = \begin{cases} HW0_{FQhi}, & F\text{=LL and CE for CHIS 2015} \\ HW0_{FQhi} \times 0.5, & F\text{=SV, SK, and SJ for CHIS 2015} \\ HW0_{FQhi}, & F\text{=LL, CE, and ABS for CHIS 2016} \end{cases} \quad (3.2)$$

3.3 Combining Samples across Quarter and Frame

Multiple independent samples were selected for each quarter and year of CHIS 2015-2016. Independent quarterly samples were selected to ensure information on the sampling frames was current.

The LL and surname samples were combined through a poststratification adjustment. The poststratification adjustment was calculated within the design strata as:

$$HW2_{Fhi} = \begin{cases} HW1_{QFhi} \times \dfrac{N_{LLh}}{\sum_Q \sum_F \sum_{i \in s_h} HW0_{QFhi}}, & F\text{=LL, SV, SK, and SJ} \\ HW1_{QFhi} \times \dfrac{N_{CEh}}{\sum_Q \sum_{i \in s_h} HW0_{QFhi}}, & F\text{=CE} \\ HW1_{QFhi}, & F\text{=ABS (CHIS 2016)} \end{cases} \quad (3.3)$$

where Q is the data collection quarter; F is the sampling frame; and s_h is the sample within design stratum h. For each year of CHIS, we used the number of end of year sampling frame counts by stratum and phone type—landline (N_{LLh}) and cell (N_{CEh})— as the poststratification totals.

3.4 Phase-2 Subsampling Adjustment

A new responsive design component was introduced with CHIS 2015-2016 for the non-ABS samples. Referred to as a nonresponse follow-up (NRFU) or the second phase of a two-phase design, a random subsample of known (or possibly) eligible telephone numbers who have yet to respond in the first phase were re-contacted using a modified recruitment protocol. Consequently, we designate the response status for the sample within each phase of the design (Table 3-1).

Table 3-1. Final response status based on phase-specific information

Response Status		
Final	Phase 1	Phase 2
Eligible Respondent	Eligible Respondent	*N/A* [2]
	Eligible Nonrespondent	Eligible Respondent
	Unknown	Eligible Respondent
Eligible Nonrespondent	Eligible Nonrespondent	Eligible Nonrespondent
	Unknown	Eligible Nonrespondent
Ineligible	Ineligible	*N/A* [2]
	Unknown	Ineligible
Unknown	Unknown	Unknown
not sampled [1]	Eligible Nonrespondent	*not sampled*
	Unknown	*not sampled*

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] *not sampled* = telephone numbers not sampled for Phase 2 and removed from the file after applying a weight adjustment

[2] Phase-1 sample members not eligible for Phase 2 and therefore are classified as not applicable (*N/A*).

Additionally, we classified sample for this adjustment as either a landline or a cell phone number based on the sampling frame, pre-screening information, or the type of phone used in our initial contact with the household. For convenience and ease of discussion, we refer to the combined LL, SV, SK, and SJ samples as the *augmented landline sample* (LS), and the combined cell phone and landline numbers ported to a cell phone as the *cell/ported sample* (CP) indexed by *T*. Note that all sampled landline numbers ported to a cell phone were designated as cell phone numbers from this point in the process and on.

A stratified Phase-2 sample was selected from the LS and CP Phase-1 samples from four strata:

- Screener Not Complete, No Contact
- Screener Not Complete, Some Contact
- Screener Complete, Adult-Only Household
- Screener Complete, Child/Teen Household

Design stratum was used as a sorting variable prior to drawing the sample to ensure geographic representation across the California counties. The following weight adjustment reflects the Phase-2 subsampling:

$$AH3_{Thki} = \begin{cases} 1, & \text{Phase-1 respondent or study ineligible (T=LS, CP)} \\ N_{Thk} / n_{Thk}, & \text{Phase 2 frame, sampled (T=LS, CP)} \\ 0, & \text{Phase 2 frame, not sampled (T=LS, CP)} \\ 1, & \text{ABS sample (T=ABS, CHIS 2016)} \end{cases} \tag{3.4}$$

where T indicates the interview type (LS, CP, or ABS); k (k=1–4) indicates the Phase-2 stratum discussed above; N_{Thk} is the number of sampled telephone numbers eligible for Phase-2 sampling by interview type and strata; and n_{Thk} is the number selected. Note that $AH2_{Thk}$=1 indicates sample that was not eligible for Phase-2 selection, including the entire ABS sample. We adjusted n_{Thk} and N_{Thk} to reflect sample either randomly chosen and not released for data collection, or released for data collection but never dialed. This adjustment was then applied to the weight in (3.3) to form:

$$HW3_{Thi} = HW2_{Fhi} \times AH3_{Thki} \tag{3.5}$$

As shown in (3.4), sampled telephone numbers eligible but not selected for Phase 2 were effectively removed from the weighting process at this stage since their adjustment $AH3_{Thki}$ is zero. The weights for the Phase-2 sample were inflated to account for the subsampling, much in the vein of Phase 1 sample having a (positive) weight to account for all those not included in the original CHIS sample.

3.5 Unknown Study Eligibility

Telephone numbers and addresses were designated as ineligible for CHIS if they met any of the following criteria:

- Non-residential (e.g., business/government, fax/modem, active cell numbers not assigned)
- Ineligible residence (e.g., all residents/cell phone owner less than 18 years of age, institutionalized residence, group quarters, 9 or more unrelated persons)
- Non-working numbers (telephone samples only)

Multiple attempts were made to determine the eligibility status of the sample in both phases of the design. Other ineligible telephone numbers were identified during data collection. For example, we

reassigned (logically imputed) cell phone numbers without a working voicemail inbox from "non-contact" to "nonworking" since this suggested that the phone number had not been activated.

There remained, however, a set of cases without a known (or imputed) eligibility status. Weights for sample with unknown residential status were distributed across the sample cases with known residential status using a weighting class adjustment similar in form to (3.3). Namely,

$$HW4_{Thi} = \begin{cases} HW3_{Thi} \times \left(\dfrac{\sum_{i \in s_g} HW3_{Thi}}{\sum_{i \in s_g} \left[I_i\left(status\ known\right) \times HW3_{Thi} \right]} \right), & \text{Phones with known residential status} \\ 0, & \text{Phones with unknown residential status} \\ HW3_{Thi}, & \text{ABS sampled addresses} \end{cases} \tag{3.6}$$

for T=LS, CP or ABS; $I_i(status\ known)$=1 if the residential status for the sampled phone number i was known (eligible or ineligible); and $I_i(status\ known)$=0 otherwise within a total of G weighting classes (g=1, …, G) defined by the cross of interview type (T=LS, CP, or ABS) and design stratum (h).

A second adjustment of similar form addressed cases known (imputed) to be residential but the within-residence eligibility status was not fully known:

$$HW5_{Thi} = \begin{cases} HW4_{Thi} \times \left(\dfrac{\sum_{i \in s_g} HW4_{Thi}}{\sum_{i \in s_g} \left[I_i\left(HHstatus\right) \times HW4_{Thi} \right]} \right), & \text{Residence with known eligibility status} \\ 0, & \text{Residence with unknown eligibility status} \\ HW4_{Thi}, & \text{Non-residence (phone or address)} \end{cases} \tag{3.7}$$

where T=LS, CP or ABS; $I_i(HHstatus)$=1 if the within-household eligibility status for the residence was known (eligible or ineligible); and $I_i(HHstatus)$=0 otherwise. This adjustment was applied within the sample weighting classes defined for (3.6). Note that only ineligible sample cases that are residences were included in this adjustment; non-residential phone numbers were excluded from the adjustment.

3.6 Screener Nonresponse Adjustment

The next adjustment in the series to finalize the household-level weight addressed CHIS screener nonresponse among those cases known or imputed to be eligible. In other words, sample cases eligible but not sampled for Phase 2, those known to be ineligible, and those with an unknown eligibility status were removed from the weighting process at this point.

We investigated candidate model covariates and interactions using CART, such as presence of children in the household and matched address for use in sending pre-contact information. However, because so little was known about nonresponding households only design stratum and frame could be used in the final model. The resulting nonresponse-adjusted weight was defined as

$$HW6_{Thi} = \begin{cases} HW5_{Thi} \times AH6_{Thi} & \text{Screener respondent} \\ 0, & \text{Screener nonrespondent} \end{cases} \tag{3.8}$$

where $HW5_{Thi}$ is defined in (3.7) and $AH6_{Thi}$ is the derived model-based nonresponse adjustment.

3.7 One-Year Household Weight

The resulting weight, $HW6_{Thi}$ is defined in (3.8), was subjected to several quality checks (Section 2.5). Based on the results, no additional adjustments were applied in either year of the study. Therefore, the final one-year household weight has the form by interview type T (LS, CP, or ABS):

$$HHW_{Thi} = HW6_{Thi} \tag{3.9}$$

4. ADULT WEIGHTING

The second stage of selection for CHIS 2015-2016 was person(s) within household, depending on the composition of the household. Below, we detail the approach used to calculate an analysis weight for analyzing responses obtained during the CHIS adult interview. Specifically, we define the initial base weights for the randomly selected adult within the household in Section 4.1. We discuss the adjustment applied to the base weights for those completing the interview in the first design phase within Section 4.2. Nonresponse to the adult interview request is addressed next (Section 4.3). Calibration adjustments to population control totals were then applied as shown in Section 4.4. A composite factor is introduced in Section 4.5 to combine dual users (landline and cell) selected from either the landline or cell sampling frames. In Section 4.6, we describe a second composite factor needed for CHIS 2016 only that when applied combines the ABS and landline/cell samples within Imperial County. The weights for the entire sample were then calibrated to estimated population projections (Section 4.7). The final one-year adult analysis weight is shown in Section 4.8. The corresponding weight for analyzing adult responses in the CHIS 2015-2016 two-year file is discussed in Section 4.9. Statistics for the adjustments and the final adult weights are provided in Appendix B.

4.1 Base Weights

One eligible adult was chosen from the sampled household. For landline telephone numbers, one adult was selected with equal probability from all those residing in the household. The adult answering the sampled cell phone was assumed to be the sole owner and therefore selected with probability one. Additional information on the sample selection procedure is provided in Section 1. As a result, the j^{th} adult base weight was defined as

$$AW1_{Thij} = HHW_{Thi} \times AA1_{Thij} \tag{4.1}$$

a function of the final household weight, HHW_{Thi} given in (3.9), and the within-household adult selection weight calculated as:

$$AA1_{Thij} = \begin{cases} ACNT_{Thij}, & T=\text{LS, ABS} \\ 1, & T=\text{CP} \end{cases} \tag{4.2}$$

As noted in report 1, a slightly different adult selection algorithm was implemented with the type of phone used for the interview. One adult was selected from all adult residents within the household ($ACNT_{Thij}$) for the landline and ABS sample cases, while the eligible adult answering a cell phone was

selected with probability one. Prior to the calculation, the range of $ACNT_{Thij}$ values was examined; values greater than three were truncated to this upper bound to limit the variation in the resulting base weight.

4.2 Phase 2 Adjustment

Section 3.4 detailed the adjustment applied to the household-level weight to account for the multiphase CHIS design introduced in 2015. Households lacking the requisite number of interviews (e.g., three interviews in a household with both an eligible child and teen), telephone numbers without a known eligibility status, and other eligible nonrespondents were eligible for Phase 2. To account for adult interviews in Phase-2 households that were actually completed in Phase 1, we apply a factor to remove the Phase-2 subsampling from the base weights in (4.1). Namely,

$$AA2_{Thik} = \begin{cases} n_{Thk} / N_{Thk}, & \text{Selected for Phase 2 but adult interview completed in Phase 1 (T=LS, CP)} \\ 1, & \text{ABS sample (T=ABS)} \\ 1, & \text{Otherwise (T=LS, CP)} \end{cases} \tag{4.3}$$

where T indicates the interview type (LS, CP, or ABS); h the design stratum; and k the Phase-2 stratum; N_{Thk} is the number of sample cases eligible for Phase-2 sampling; and n_{Thk} is the number selected for Phase 2. The resulting adjusted weight was calculated as

$$AW2_{Thij} = AW1_{Thij} \times AA2_{Thik} \tag{4.4}$$

4.3 Adult Nonresponse Adjustment

As with the CHIS screener, repeated attempts were made to obtain a completed adult interview. Sampled adults with partially completed interviews with responses through Section K were classified as respondents and nonrespondents otherwise. The nonresponse-adjusted weight was calculated as:

$$AW3_{Thij} = \begin{cases} AW2_{Thij} \times AA3_{Thij} & \text{Respondents} \\ 0, & \text{Nonrespondents} \end{cases} \tag{4.5}$$

where T indicates the interview type (LS, CP, or ABS) and $AW2_{Thij}$ is defined in (4.4). The multiplier $AA3_{Thij}$ is the nonresponse adjustment derived using much of the same procedures as discussed for the screener nonresponse-adjusted weight (Section 3.6). Specifically, we implemented a CART analysis to determine a parsimonious set of variables to calculate a model-based nonresponse adjustment.

SUDAAN's PROC WTADJUST was used to form $AA3_{Thij}$ in (4.5). In addition to the nonresponse model covariates are provided in Table 4-1, the following variables were evaluated but not included in the adjustment model: use of Medi-Cal by the household members, presence of mailing address, and child-first household (see Section 5.1 for a detailed definition).

Table 4-1. Variables for nonresponse adjustment applied to the adult weights

Level	Description
Geographic/Sampling	▪ Design stratum / county ▪ Phase 2 sample member
Household	▪ Number of adults (1, 2, and 3 or more) ▪ Address available with sampled number (landline only) ▪ Likely Hispanic surname (landline only) ▪ Likely Asian surname (landline only) ▪ Number of children 0-17 years old ▪ Call answered on cell phone
Person	▪ Selected adult has medical conditions

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

4.4 Calibration Adjustment to NHIS

The overall sum of $AW4_{Thij}$ (4.7) for the telephone samples (T=LS, CP) by interview status estimates the total adults in the population of landline users, cell phone users, and dual users in California. However, misalignment of the estimates with reality may occur since the mutually exclusive "conceptual" frames, one for each phone-use status, does not exist to draw our samples. Thus, a calibration adjustment was implemented to align the weight sums for the four groups generated from the *telephone* samples: landline only (L), cell phone only (C), dual-use landline sample (DL), and dual-use cell sample (DC). As shown below, this adjustment for the CHIS 2016 ABS sample was set to one.

As in past rounds of CHIS, we used estimates from the National Health Interview Survey (NHIS) by phone usage-status (landline only, cell only, dual use) as the source for the population control totals (Blumberg & Luke, 2016). The calibrated weights were calculated as:

$$AW4_{Thij} = \begin{cases} AW3_{Thij} & \text{ABS sample (T=ABS)} \\ AW3_{Thij} \times AA4_{Lhij} & \text{landline-only users} \\ AW3_{Thij} \times AA4_{DLhij} & \text{dual-use landline sample} \\ AW3_{Thij} \times AA4_{DChij} & \text{dual-use cell sample} \\ AW3_{Thij} \times AA4_{Chij} & \text{cell phone-only users} \end{cases} \quad (4.6)$$

where the model-based adjustment factors for landline-only users ($AA4_{Lhij}$), cell phone-only users ($AA4_{Chij}$), and dual users ($AA4_{DLhij}$ and $AA4_{DChij}$) calculated with SUDAAN's PROC WTADJUST using NHIS control totals. As shown in (4.6), all ABS sample cases were excluded from this adjustment and $AW3_{Thij}$ (4.5) was carried forward unchanged. Note that WTADJUST enables the calculation of adjustment for either nonresponse or calibration (see, e.g., Kott, 2006). Additional information for the calibration adjustment is provided in Section 7.3.4.

4.5 Composite of Cell Phone and Landline Samples

The calibrated survey weights for our CHIS telephone-sample respondents now aligned with population estimates (controls) generated from a large national survey for two of the four CHIS groups—landline only and cell only. Summing the weights for the dual users, however, would estimate twice the size of the dual-use population. The next step was to adjust (composite) the weights for the dual-use telephone interviews so that when combined they would reproduce the population size regardless of their frame source.

The composite weight was constructed as:

$$AW5_{Thij} = \begin{cases} AW4_{Thij} & \text{landline-only users (T=LS)} \\ AW4_{Thij} \times \lambda_{A1} & \text{dual users, landline frame (T=LS, CP)} \\ AW4_{Thij} \times (1-\lambda_{A1}) & \text{dual users, cell frame (T=LS, CP)} \\ AW4_{Thij} & \text{cell phone-only users (T=CP)} \\ AW4_{Thij} & \text{ABS sample (T=ABS)} \end{cases} \quad (4.7)$$

where $0 \leq \lambda_{A1} \leq 1$ is the composite factor for the adult dual-use telephone-sample respondents. Various methods were examined to determine λ_{A1}, such as setting λ_{A1}=0.5 for all dual-use landline respondents to give equal weight for landline and cell phone usage and λ_{A1} equal to the relative sample size within reported stratum. The resulting composite factor was the latter option, i.e., the proportion of dual-use landline sample out of the combined dual-use landline and cell samples within by design stratum. This methodology is applied in several other surveys (see, e.g., Levine and Harter, 2015) and demonstrated lower CHIS standard errors for a small set of variables included in the evaluation in comparison with the λ_{A1}=0.5 procedure.

4.6 Composite of Cell/Landline and Northern Imperial County Samples

Next, we combined the CHIS 2016 respondents from the combined landline/cell samples with those from the ABS sample using a second composite factor for residents of Imperial County only. The composite weight was defined as

$$AW6_{hij} = \begin{cases} AW5_{Thij} \times \lambda_{A2} & \text{landline/cell respondents in Northern Imperial County (CHIS 2016)} \\ AW5_{Thij} \times (1-\lambda_{A2}) & \text{ABS sample (CHIS 2016)} \\ AW5_{Thij} & \text{Otherwise} \end{cases} \quad (4.8)$$

where $0 \leq \lambda_{A2} \leq 1$ is the composite factor for the Imperial County ABS sample. As with (4.7), the lambda value for this compositing step was calculated using the relative sample sizes.

4.7 Calibration Adjustment to Department of Finance Projections

In keeping with the specified statement of work for CHIS 2015-2016, RTI calibrated the composite weights, $AW6_{hij}$ in (4.7), to adjusted values of population projections supplied by the State of California's Department of Finance. Population estimates associated with California residents living in group quarters (e.g., nursing homes, prisons) and others who were not eligible for CHIS was estimated and excluded from the population controls, using techniques like those documented in the *CHIS 2013-2014 Methodology Series: Report 5 – Weighting and Variance Estimation*.[3] The calibrated weight was calculated as,

$$AW7_{hij} = AW6_{hij} \times AA7_{hij} \quad (4.9)$$

as a function of the composite weight $AW6_{hij}$ (4.8) and a calibration adjustment factor $AA6_{hij}$ obtained from SUDAAN PROC WTADJUST. Calibration variables, calculation of the estimated calibration control totals, and information associated with the calibration procedure are detailed in Chapter 7. The model covariates and interactions mirrored those used in prior rounds of CHIS (see Section 7.2).

For Imperial County, we designated the telephone-sample respondents either as falling within the targeted area identified for the ABS sample, or residing outside this area but within the county boundary.

[3] http://healthpolicy.ucla.edu/chis/design/Documents/chis2013-2014-method-5_2017-01-11.pdf

The population quantities for targeted area (Yes vs. No) were estimated and used in the calibration adjustment.

4.8 Adult One-Year Analysis Weight

The resulting adult weights, $AW7_{hij}$ in (4.9), were evaluated for outlier values. As noted in Appendix B, a few weights were trimmed. Therefore, the final one-year adult weight was constructed as follows:

$$ADW_{hij} = AW7_{hij} \times AA8_{hij} \tag{4.10}$$

where $AA8_{hij}$ is the adjustment to minimize the size of the outlier weights.

4.9 Adult Two-Year Analysis Weight

Two-year weights were created for analyzing the CHIS 2015-2016 combined data files. This process involved combining the two one-year weights within two steps. First, composite factors were derived as the relative sample size by year and reported stratum, similar to the calculation discussed in Section 4.5. A two-year composite weight for the adult respondents was constructed as:

$$AW2Y1_{hij} = \begin{cases} ADW_{hij} \times \lambda_{2015h} & \text{2015 respondents} \\ ADW_{hij} \times (1 - \lambda_{2015h}) & \text{2016 respondents} \end{cases} \tag{4.11}$$

where λ_{2015h} is the proportion of adults who responded in 2015 by reported stratum h, and ADW_{hij} is the one-year analysis weight shown in equation (4.10).

Then, in keeping with past rounds of CHIS, we calibrated the two-year weights to the Department of Finance adjusted population estimates used for the CHIS 2016 weights. Calibration variables and information associated with the calibration procedure are detailed in Chapter 7. The two-year adult calibrated weight was calculated as,

$$ADW2_{hij} = AW2Y1_{hij} \times AA2Y2_{hij} \tag{4.12}$$

as a function of the composite weight $AW2Y1_{hij}$ (4.11) and a calibration adjustment factor $AA2Y2_{hij}$ obtained from SUDAAN PROC WTADJUST. No additional adjustments were applied after the final calibration step.

5. CHILD WEIGHTING

Children, ages 11 years and younger, of the randomly chosen adult in households participating in CHIS were also eligible for the study. Information on the children and interview responses were collected either from the adult participant or, if relevant, from the other legal parent who completed the screener.

Below, we describe how the child (proxy interview) analysis weight were calculated. Many of the weighting steps follow those discussed for the adult weights. Specifically, we define the input values for the child weights in Section 5.1 that were then adjusted to account for the child-level sampling (Section 5.2). We briefly describe the nonresponse adjustment applied to the weights in Section 5.3, followed by an initial calibration adjustment to account for phone usage in Section 5.4. A composite factor is discussed in Section 5.5 to combine dual users (landline and cell) selected from the landline or the cell sampling frames like the method discussed for the adult weight. A second composite adjustment was implemented to incorporate the Imperial County oversample (Section 5.6). These weights were then calibrated to population projections (Section 5.7). The child one-year analysis is shown in Section 5.8 that may include additional refinements. The corresponding weight for analyzing child (proxy) responses in the CHIS 2015-2016 two-year file is discussed in Section 5.9. Statistics for the adjustments and the final child weights are provided in Appendix B.

5.1 Adjustment for Adult Nonresponse

Households with children were classified into two groups based on whether the child-first methodology was used. A child-first household was one where the screener respondent was not the selected adult participant but they were the legal guardian of the adult participant's children. A household with children not employing the child-first methodology (a child-second household) was one where the screener respondent was either the selected adult participant or not a legal guardian of the adult participant's children. Child selection occurred in the screener for the child-first households. By contrast, selection of a child occurred in *Section G* of the adult questionnaire for the other child households; thus, sampling only occurs for these cases if the adult completes *Section G*. Note that the child-first methodology was only available for households contacted via a landline number or ABS household because the screener respondent and adult participant were automatically one in the same for cell phone sample.

The input values for the child weights were linked to the probability of selection for their parent. Consequently, the input values for the child weights were adjusted to account for Section G adult for the child-second households. Specifically, the adult-level input weights were constructed as follows:

$$AW9_{Thij} = \begin{cases} AW2_{Thij}, & \text{child-first household} \\ AW2_{Thij} \times AA9_{Thij}, & \text{child-second household and adult responds} \\ 0, & \text{no children, or child-second household with no adult response} \end{cases} \quad (5.1)$$

a function of the adult weight, $AW2_{Thij}$ in (4.4), that was adjusted for cases not subsampled for Phase 2 and a new nonresponse adjustment, $AA9_{Thij}$ that accounted for missing Section G responses. The adjustment was calculated by design strata only for child-second households as a ratio of the $AW2_{Thij}$ weight sums divided by the weights sums for those with child roster information. This approach was like the unknown residential status adjustment shown in (3.6).

5.2 Base Weights

The child-level base weights, conditional on the selection of the parent, were calculated as the inverse selection probability by age group. Let n_{1ij} represent the number of eligible children ages 0-5 years of the adult respondent and n_{2ij} represent the number of children ages 6-11 years of the adult respondent within household i. We define the probability of selection as follows:

$$CA0_{Thij} = \begin{cases} 2n_{1Thij} \big/ \left(2n_{1Thij} + n_{2Thij}\right), & \text{children ages 0-5 years} \\ n_{2Thij} \big/ \left(2n_{1Thij} + n_{2Thij}\right), & \text{children ages 6-11 years} \end{cases} \quad (5.2)$$

Note that children ages 0-5 years have twice the likelihood of being selected for the study by design. The resulting child-level base weight was defined as

$$CW0_{Thij} = \begin{cases} AW9_{Thij} / CA0_{Thij}, & \text{household with one parent} \\ \left(AW9_{Thij} / CA0_{Thij}\right) \times 0.5, & \text{household with two parents} \end{cases} \quad (5.3)$$

with terms defined above. For households containing two parents of the selected child, we adjust the weight to account for twice the probability of selection.

5.3 Nonresponse Adjustment

Prior to all nonresponse adjustments, we evaluated data for differential patterns of refusals using correlational statistics and CART. Nonresponse for the child interview could have occurred at two points during recruitment—parental permission and the interview. Because of minimal differences in the CART models for the two nonrespondent groups, only one nonresponse model adjustment was used of the form:

$$CW2_{Thij} = \begin{cases} CW1_{Thij} \times CA2_{Thij}, & \text{child-interview respondents} \\ 0, & \text{child-interview nonrespondents} \end{cases} \quad (5.4)$$

where $CA2_{Thij}$ is the nonresponse adjustment calculated using SUDAAN's PROC WTADJUST. The nonresponse model included design stratum, frame type, adult count, whether it was a child-first household, the number of youth under 18, whether there was an address available with the sampled telephone number (landline only), call answered on cell phone, and anyone in the household covered by Medi-Cal.

5.4 Calibration Adjustment to NHIS

We calibrated the child-level weights by year to the NHIS phone-usage estimates using the same methodology as implemented for the adult weight adjustment (Section 4.4) for consistency. The calibrated weights were calculated as:

$$CW3_{Thij} = \begin{cases} CW2_{Thij} & \text{ABS sample (T=ABS)} \\ CW2_{Thij} \times CA3_{Lhij} & \text{landline-only household} \\ CW2_{Thij} \times CA3_{DLhij} & \text{dual-use landline sample} \\ CW2_{Thij} \times CA3_{DChij} & \text{dual-use cell sample} \\ CW2_{Thij} \times CA3_{Chij} & \text{cell phone-only household} \end{cases} \quad (5.5)$$

where the adjustment factors for landline-only users ($CA3_{Lhij}$), cell phone-only users ($CA3_{Chij}$), and dual users ($CA3_{DLhij}$ and $CA3_{DChij}$) were calculated with SUDAAN's PROC WTADJUST. No adjustment was calculated for the ABS sample.

5.5 Composite of Cell Phone and Landline Samples

As with the adult weight, the calibrated weights in (5.5) for the dual-use households were twice the size of the population and required adjustment. The resulting composite weight by year was constructed as:

$$CW4_{hij} = \begin{cases} CW3_{Thij} & \text{landline-only users} \\ CW3_{Thij} \times \lambda_{C1} & \text{dual users, landline frame} \\ CW3_{Thij} \times (1-\lambda_{C1}) & \text{dual users, cell frame} \\ CW3_{Thij} & \text{cell phone-only users} \\ \text{CW3}_{Thij} & \text{ABS sample} \end{cases} \quad (5.6)$$

where $0 \leq \lambda_{C1} \leq 1$ was the composite factor for the child (proxy) dual-use respondents. For consistency, the same lambda values used for the adult weights were applied to the child weights.

5.6 Composite of Cell/Landline and Northern Imperial County Samples

Next, we created a single weight to account for the combined landline/cell samples and those from the ABS sample using a second composite factor. The composite weight was calculated as:

$$CW5_{hij} = \begin{cases} CW4_{Thij} \times \lambda_{C2} & \text{landline/cell respondents in Imperial County ABS area (CHIS 2016)} \\ CW4_{Thij} \times (1-\lambda_{C2}) & \text{ABS sample (CHIS 2016)} \\ CW4_{Thij} & \text{Otherwise} \end{cases} \quad (5.7)$$

where $0 \leq \lambda_{C2} \leq 1$ is the composite factor for the Imperial County oversample area, reflecting the relative sample sizes from the landline/cell sample as with the first composite factor (5.6).

5.7 Calibration Adjustment to Department of Finance Projections

The composite weights, $CW5_{hij}$ in (5.7), were calibrated to the eligible population projections (sans CHIS-ineligible residences) supplied by the State of California's Department of Finance. The calibrated weight was calculated as

$$CW6_{hij} = CW5_{hij} \times CA6_{hij} \quad (5.8)$$

a function of the composite weight and a calibration adjustment factor $CA6_{hij}$ obtained from SUDAAN PROC WTADJUST. Calibration variables, calculation of the estimated population values used in calibration, and information associated with the calibration procedure are detailed in Chapter 7. Model covariates mirrored those used for raking in prior rounds of CHIS in addition to variables identified through subsequent analyses.

5.8 Child One-Year Analysis Weight

The calibrated child weights, $CW6_{hij}$ in (5.8), were evaluated for outlier values. As noted in Appendix B, a few weights were trimmed. The final one-year child weight was constructed as follows:

$$CHW_{hij} = CW6_{hij} \times CA7_{hij} \tag{5.9}$$

where $CA7_{hij}$ was the adjustment to minimize the size of the outlier weights.

5.9 Child Two-Year Analysis Weight

Two-year weights were created for analyzing the CHIS 2015-2016 combined child interview files. This process followed the steps outlined for the adult weights in Section 4.9 to maintain consistency of the weights. Details of the calibration control totals are found in Chapter 7 of this report.

6. TEEN WEIGHTING

Teenaged children, ages 12 to 17, of the randomly chosen adult were eligible for the study. In contrast to the child (proxy) interview, one randomly chosen teen was recruited to conduct an interview only after receiving permission from a parent.

Below, we describe our approach calculating a teen analysis weight for analyzing an annual CHIS data file. Steps to calculate the teen weights follow many of those specified for the child weight. Specifically, we define the teen base weight in Section 6.1. We describe in Section 6.2 nonresponse adjustments applied to the weights. This discussion is followed by one for a calibration adjustment to population control totals for phone usage (Section 6.3). The composite factor, like the one discussed for the child weight, is outlined in Section 6.4 to combine dual users (landline and cell) selected from both telephone frames. A second composite adjustment for including the Imperial County ABS areas is described in Section 6.5. We constructed a calibration adjustment to population projections (Section 6.6) and with a few refinements created the final teen one-year weight (Section 6.7). The weight for analyzing teen interviews within the CHIS 2015-2016 two-year file is discussed in Section 6.8. Statistics for the adjustments and the final teen weights are provided in Appendix B.

6.1 Base Weights

The teen base weights, conditional on the selection of the parent, were calculated as the inverse selection probability. Unlike sampling for children, teens were selected with equal probability. The resulting teen base weight was calculated as

$$TW0_{Thij} = \begin{cases} AW2_{Thij} \times TCNT_{Thij}, & \text{household with one parent} \\ AW2_{Thij} \times TCNT_{Thij} \times 0.5, & \text{household with two parents} \end{cases} \quad (6.1)$$

a function of the adult base weight, $AW2_{Thij}$ in (4.4) and $TCNT_{Thij}$ defining the number of eligible teens linked to the selected adult. For households containing two parents of the selected teen, the weight was adjusted to account for twice the probability of selection.

6.2 Adjustment for Teen Nonresponse

Nonresponse could have occurred at two points during recruitment of the teen for CHIS. The first was parental permission and the second was from the selected teen. However, because of small sample sizes in the nonresponse groups, only one adjustment was applied to the weights. Additionally,

insufficient information was available to identify relevant covariates via CART for a model-based solution. Thus, the adjustment for nonresponding teens was based on sample frame, stratum, and indicator for child-first household of the form:

$$TW1_{Thij} = \begin{cases} TW0_{Thij} \times TA1_{Thij}, & \text{responding teen} \\ 0, & \text{nonresponding teen} \end{cases} \tag{6.2}$$

where $TA1_{Thij}$ is the nonresponse adjustment calculated using SUDAAN's PROC WTADJUST with model covariates identified through a CART analysis.

6.3 Calibration Adjustment to NHIS

The nonresponse-adjusted teen weights were calibrated to phone-usage estimates for California using the same methodology as implemented for the adult weight adjustment (Section 4.4). The calibrated weights were constructed as:

$$TW2_{Thij} = \begin{cases} TW1_{Thij} & \text{ABS sample (T=ABS)} \\ TW1_{Thij} \times TA2_{Lhij} & \text{landline-only household} \\ TW1_{Thij} \times TA2_{DLhij} & \text{dual-use landline sample} \\ TW1_{Thij} \times TA2_{DChij} & \text{dual-use cell sample} \\ TW1_{Thij} \times TA2_{Chij} & \text{cell phone-only household} \end{cases} \tag{6.3}$$

where the adjustment factors for landline-only users ($TA4_{Lhij}$), cell phone-only users ($TA4_{Chij}$), and dual users ($TA4_{DLhij}$ and $TA4_{DChij}$) were calculated with SUDAAN's PROC WTADJUST.

6.4 Composite of Cell Phone and Landline Samples

The composite weight, calculated to "down weight" the dual-use estimates generated with $TW3_{ij}$, were constructed as:

$$TW3_{Thij} = \begin{cases} TW2_{Thij} & \text{landline-only users} \\ TW2_{Thij} \times \lambda_{T1} & \text{dual users, landline frame} \\ TW2_{Thij} \times (1-\lambda_{T1}) & \text{dual users, cell frame} \\ TW2_{Thij} & \text{cell phone-only users} \\ TW2_{Thij} & \text{ABS sample (T=ABS)} \end{cases} \tag{6.4}$$

where $0 \le \lambda_{T1} \le 1$ is the composite factor for the teen respondents in the dual-use (landline and cell phone) households. Similar to the child weights the same composite factor derived for the adult weights, varying by self-reported stratum, was used for teens.

6.5 Composite of Cell/Landline and Northern Imperial County Samples

Next, we "down weighted" respondents within the Imperial County ABS area obtained from the combined landline/cell samples and ABS frame using a second composite factor. The composite weight was calculated as:

$$TW4_{hij} = \begin{cases} TW3_{Thij} \times \lambda_{T2} & \text{landline/cell respondents in Northern Imperial County (T=LS, CP)} \\ TW3_{Thij} \times (1-\lambda_{T2}) & \text{ABS sample (T=ABS)} \\ TW3_{Thij} & \text{Otherwise} \end{cases} \tag{6.5}$$

where $0 \le \lambda_{T2} \le 1$ is the composite factor for the Imperial County oversample area, reflecting the relative sample sizes from the landline/cell sample as with the first composite factor (6.4).

6.6 Calibration Adjustment to Department of Finance Projections

The composite weights were calibrated to the eligible population projections supplied by the State of California's Department of Finance as discussed for the adult and child weights. We calculated the calibrated weight as

$$TW5_{hij} = TW4_{hij} \times TA5_{hij} \tag{6.6}$$

a function of the composite weight and a calibration adjustment factor $TA5_{hij}$ obtained from SUDAAN PROC WTADJUST. Calibration variables, calibration values, and information associated with the calibration procedure are detailed in Chapter 7. Model covariates mirrored those used in prior rounds of CHIS in addition to variables identified through subsequent analyses.

6.7 Teen One-Year Analysis Weight

The calibrated weights, $TW5_{hij}$ in (6.6), were evaluated for outlier values. As noted in Appendix B, a few weights were trimmed. The final one-year child weight was constructed as follows:

$$TNW_{hij} = TW5_{hij} \times TA6_{hij} \tag{6.7}$$

where $TA6_{hij}$ was the adjustment to minimize the size of the outlier weights.

6.8 Teen Two-Year Analysis Weight

Two-year weights were created for analyzing teen responses within the CHIS 2015-2016 combined analysis files. Procedures for producing the two-year weights followed those outlined for the adult weights in Section 4.9 to maintain consistency of the weights. Details of the calibration control totals are found in Chapter 7 of this report.

7. CALIBRATION CONTROL TOTALS

Calibration to population values is an important attribute of the CHIS weights. Section 7.1 contains an overview of weight calibration and highlights the many benefits of such efforts. Section 7.2 contains the dimensions used in the final calibration models, along with steps to address small sample size for certain dimensions. Population sources accessed for key information are detailed in Section 7.3. Steps to convert the population information into usable calibration control totals are discussed in Section 7.4.

7.1 Calibration Procedure

Calibration is a weight adjustment method where survey-estimated population counts are constrained to equal their corresponding population control totals. If the population characteristics are associated with a survey characteristic, then the estimated characteristic will have a smaller standard error with calibration compared to its size with unadjusted analysis weights (Kott, 2006; Valliant et al., 2013). Poststratification and raking are types of weight calibration. With poststratification, characteristics are interacted (e.g., sex crossed with levels of race/ethnicity) to form a relatively large number of weighting cells (classes). Using too many characteristics could result in cells with a small amount of sample, resulting in an increase in the variability of the weights and consequently a reduction in precision for estimates using these weights. Small cells are generally collapsed with larger cells to improve precision but the sometimes ad hoc collapsing can increase bias in the estimates (Kim et al., 2007). Raking (Kalton & Flores Cervantes, 2003), in its traditional form, only using the marginal control totals and no interactions, thereby including more covariate than poststratification but excluding finer adjustments that could benefit the survey estimates.

Calibration using the WTADJUST procedure in SUDAAN (Section 2.2.2) as with CHIS 2015-2016 combines the benefits of poststratification and raking by allowing many controls with constraints on the adjustment to control decrease in precision. Specifically, calibration allows a combination of marginal control (e.g., design strata) and interactions (e.g., region by sex by race/ethnicity). By-hand collapsing possibly needed for poststratification and raking can be addressed through the constraints. If constraints are used, then differences between the adjusted weight sums and the population controls may exist but typically this is only by a small amount.

Calibration adjustments were used twice in CHIS 2015-2016. The first was to adjust estimated counts by telephone usage (landline only, cell only, and dual use) to population estimates for the state of

California. Information for the adult, child and teen adjustments are discussed in Sections 4.4, 5.4, and 6.3, respectively. The second calibration adjustment was implemented to align the weight sums to person-level estimates by several characteristics, while maintaining the phone usage adjustments in the first calibration procedure. Information for the adult, child and teen adjustments are discussed in Sections 4.7, 5.7, and 6.6, respectively. The control total used in the calibration models are detailed in the next section (Section 7.2). Because population totals required for the adjustment did not exist, needed population estimates were generated from population information that was available. The control total sources for the two calibration adjustments are listed in Section 7.3. Estimation methods for the CHIS control totals are detailed in Section 7.4.

7.2 Calibration Model Dimensions

The 13 weight calibration dimensions used in CHIS 2015-2016 are shown in Table 7-1. These dimensions followed those specified in prior years of the study to maximize continuity. Specifically, Dimensions 1-8 involved combinations of demographic characteristics (age, sex, race/ethnicity) and reported geography (county, region, state). Regions of the state are shown in Table 7-2. Note that the number of age groups is provided in parentheses, such as age groups (3) = under 12 years, 12 to 17 years, and 18 years or older shown for Dimension 1. Dimensions 9 and 10 included education of the responding adult and number of adults in the household, respectively, crossed with region. Dimension 11 is complex in nature and elaborated on below. Dimension 12 interacted age groups (3) with household phone-usage status (landline only, cell phone only, and dual user). Dimension 13, new to the CHIS series, involved the calculation of counts to incorporate the targeted area within Imperial County.

Levels within the dimensions were sometimes collapsed for situations where there were only a small number of (or no) respondents with that characteristic. Where possible, we included in the calibration model all covariate levels and interactions that were sometimes collapsed in the one-year files because of small sample sizes. Additionally, even though the Imperial County ABS oversample was included in CHIS 2016 only (Dimension 13), an indicator for living within the oversample area was created for those CHIS 2015 respondents who reported living in Imperial County. This step allowed the inclusion of this information as an extra dimension in the two-year calibration model compared with CHIS 2015.

Table 7-1. Dimensions used in weight calibration

Dimension	Description	Categories		CHIS 2015			CHIS 2016			CHIS 2015-2016		
				Adult	Child	Teen	Adult	Child	Teen	Adult	Child	Teen
1	Region[2] (7) x Age groups (3) x Sex (2)	Under 12 years, male	Under 12 years, female									
		12 to 17 years, male	12 to 17 years, female		✓	✓	✓	✓	✓	✓	✓	✓
		18 years or older, male	18 years or older, female									
2	Region[2] (7) x Age groups (9)	Under 6 years	6 to 11 years									
		12 to 17 years	18 to 24 years									
		25 to 29 years	30 to 39 years	✓	✓	✓	✓	✓		✓	✓	✓
		40 to 49 years	50 to 64 years									
		65 years or older										
3	Age groups (13) x Sex (2)	Under 4 years, male	Under 4 years, female									
		4 to 7 years, male	4 to 7 years, female									
		8 to 11 years, male	8 to 11 years, female									
		12 to 14 years, male	12 to 14 years, female									
		15 to 17 years, male	15 to 17 years, female									
		18 to 24 years, male	18 to 24 years, female									
		25 to 30 years, male	25 to 30 years, female	✓	✓	✓	✓	✓		✓	✓	✓
		31 to 37 years, male	31 to 37 years, female									
		38 to 45 years, male	38 to 45 years, female									
		46 to 53 years, male	46 to 53 years, female									
		54 to 64 years, male	54 to 64 years, female									
		65 to 77 years, male	65 to 77 years, female									
		78 years or older, male	78 years or older, female									

(continued)

Table 7-1. Dimensions used in weight calibration (continued)

Dimension	Description	Categories		CHIS 2015			CHIS 2016			CHIS 2015-2016		
				Adult	Child	Teen	Adult	Child	Teen	Adult	Child	Teen
4	SPAs (8), HSRs (6), Remainder	Remainder of CA		✓	✓	✓	✓	✓	✓	✓	✓	✓
		SPA 1 – Antelope Valley	SPA 2 – San Fernando									
		SPA 3 – San Gabriel	SPA 4 – Metro									
		SPA 5 – West	SPA 6 – South									
		SPA 7 – East	SPA 8 – South Bay									
		HR 1 – North Coastal	HR 2 – North Central									
		HR 3 – Central	HR 4 – South									
		HR 5 – East	HR 6 – North Inland									
5	Region[2] (7) x Race/ethnicity (7)	Latino	NL White	✓		✓	✓	✓		✓	✓	.
		NL African American	NL American Indian									
		NL Asian	NL Native Hawaiian									
		NL Two or more races										
6	Race/ethnicity (7) x Gender (2)	Latino, Male	Latino, Female	✓		✓	✓		✓	✓	✓	✓
		NL White, Male	NL White, Female									
		NL African American, Male	NL African American, Female									
		NL American Indian, Male	NL American Indian, Female									
		NL Asian, Male	NL Asian, Female									
		NL Native Hawaiian, Male	NL Native Hawaiian, Female									
		NL Two or more races, Male	NL Two or more races, Female									
7	Asian groups (6)	NL Chinese only	NL Korean only	✓	✓	✓	✓	✓	✓	✓	✓	✓
		NL Filipino only	NL Vietnamese only									
		Other or non-Asian only	NL Japanese only									

(continued)

Table 7-1. Dimensions used in weight calibration (continued)

Dimension	Description	Categories		CHIS 2015			CHIS 2016			CHIS 2015-2016		
				Adult	Child	Teen	Adult	Child	Teen	Adult	Child	Teen
8	Reported stratum[1] (44) x Race/ethnicity (3)	Latino		✓	✓		✓			✓	✓	.
		NL White										
		NL Non-White										
9	Region[2] (7) x Education[3] (4)	Not applicable (age < 18 years)	Less than High School	✓	✓	✓	✓	✓	✓	✓		✓
		High School grad or GED recipient	At least some college									
10	Region[2] (7) x # Adults in HH (3)	0 or 1 adult	2 adults				✓			✓		
		3 or more adults										
11	Region[2] (7) x Non-telephone[4] (16)			✓	✓	✓	✓	✓	✓	✓	✓	✓
12	Region[2] (7) x Household phone usage	Landline only	Dual-user	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Cell-only										
13	Northern Imperial County (3)	In Northern Imperial County	Remainder of Imperial County				✓	✓	✓	✓	✓	✓
		Remainder of CA										

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Note: NL = Non-Latino; 18 years+ = 18 years of age or older; SPA = Service Planning Area; HSR = Health Service Region.

[1] Interactions with reported stratum (44) were suppressed for brevity. See Table 1-1 for a description of the 44 strata.

[2] Interactions with Region (7) were suppressed for brevity. See Table 7-2 for a description of the seven regions of California.

[3] Education level for the adult was used for children and teens for the 2015 and 2016 weights; however, for 2016 teen the adjustment had to be simplified to main effect only.

[4] See Table 7-3. Dimension 11, non-telephone adjustment category definition.

Table 7-2. Regions in California (CREGION)

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, 2015-2016 California Health Interview Survey.

Dimension 11 (Table 7-3) included characteristics associated with non-telephone households identified previously for CHIS. Through calibration, the biasing effects of excluding non-telephone households from the study, estimated to be 2.5% of households in California (Blumberg and Luke, 2016), is minimized. Counts for this dimension were also estimated using the same procedures as the other dimension; these procedures are detailed in Section 7.3.

Table 7-3. Non-telephone calibration dimension 11

Level	Household tenure	Age in years	Educational attainment	Number of adults in the household
1	Own	0 to 17	*N/A*	1
2	Rent	0 to 17	*N/A*	1
3	Own	0 to 17	*N/A*	2 or more
4	Rent	0 to 17	*N/A*	2 or more
5	Own	18 to 30	Up to high school	1 or more
6	Own	31 to 64	Up to high school	1 or more
7	Own	65 and older	Up to high school	1 or more
8	Own	18 to 30	Greater than high school	1 or more
9	Own	31 to 64	Greater than high school	1 or more

(continued)

Table 7-3. Non-telephone calibration dimension 11 (continued)

Level	Household tenure	Age in years	Educational attainment	Number of adults in the household
10	Own	65 and older	Greater than high school	1 or more
11	Rent	18 to 34	Up to high school	1 or more
12	Rent	35 and older	Up to high school	1
13	Rent	35 and older	Up to high school	2 or more
14	Rent	18 to 34	Greater than high school	1 or more
15	Rent	35 and older	Greater than high school	1
16	Rent	35 and older	Greater than high school	2 or more

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

N/A = not applicable.

7.3 Sources for Population Control Totals

No individual source was available to address the calibration control total needs for CHIS. In keeping with prior rounds of the study, multiple government databases were combined to produce estimated population values used in two separate adjustments for each person-level weight—phone usage and population distributions within California. We describe the sources below.

7.3.1 California Department of Finance Population Predictions and Estimates

As in prior years of CHIS, the California Department of Finance (DOF) population projections was the primary source for calculating estimated control totals used in weight calibration. Population counts by county and person-level characteristics (Table 7-4) were provided for 2015 and 2016 for yearly file adjustments. This sole source by year produced estimates for adult, child and teen weight because projections are provided by single year of age up to 100 years. Additional information on the history of the DOF projections is provided in the *CHIS 2013-2014 Methodology Series: Report 5 – Weighting and Variance Estimation*.

The DOF projections, however, were not in perfect alignment with CHIS and additional adjustments were required. First, DOF projections followed the U.S. Office of Management and Budget (OMB) modified race definition and as shown in Table 7-4 did not include an “other race” group (OMB, 1997). With CHIS, respondents could designate one or more of five main racial categories—White,

Table 7-4. Definition of counts available in 2015 and 2016 California DOF population files

Category	Levels[1]
Age groups (101)	Age less than 1 year
	Age 1 year, …, Age 100 years or more (by single year of age)
Sex (2)	Male
	Female
Race/ethnicity (13)	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

Source: California Department of Finance projections.

[1] Population projections were provided by the cross of the three categories and county.

Black/African American, American Indian/Alaska Native, Asian, or Native Hawaiian/Other Pacific Islander. All open-end responses that could not be collapsed into a single or multi-race using this groups were classified as "other" and for the purposes of weighting were imputed as one of the OMB categories. (See discussion of OMBSRREO in Section 8.4.2)

DOF projections also included California residents who live in group quarters, a population that was ineligible for CHIS. Census 2010 files were used to estimate the proportion of persons in group quarters; these values were subtracted from the DOF projections, and these proportions were removed from the DOF estimates (see Section 7.4.1).

Additionally, the person characteristics on the DOF file did not allow the estimate of population counts for all calibration dimensions. Therefore, additional sources were required for this purpose as discussed below.

7.3.2 Census 2010 Files

As in prior years, data from the 2010 Census was used as source information for CHIS in four ways.

- The proportion of CHIS-ineligible residents living in group quarters was estimates from the 2010 Census Summary File 1 (SF1; U.S. Census Bureau, 2012a). Section 7.6.1 describes the details of this process. Information available from the SF1 is provided in Table 7-5.
- The SF1 was adjusted by information on the 2010 Census Modified Race File (U.S. Census Bureau, 2012b) to calculate population counts for the "other race" group.
- The SF1 was also used to generate population distributions by Service Planning Areas (SPAs) within Los Angeles County and by Health Service Region (HSR) within San Diego County, which were then applied to the DOF population total for that county. As with the DOF projections, these counts were adjusted to remove those estimated to live in group quarters and used for Dimension 4 (Table 7-1).
- Proportions of all Asians by Asian nationality for Dimension 7 (Table 7-1) were estimated from the 2010 Census Summary File 2 (SF2; U.S. Census Bureau, 2012c) and applied to the DOF population total for Asians.

Table 7-5. Definition of variables available on the 2010 Census Summary File [1]

Category	Levels
Stratum (44)[1]	
Sex (2)	Male Female
Age groups (3)	Less than 18 years old 18-64 years old 65 years old or older
Ethnicity (3)	Latino Non-Latino, White alone Other
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

Source: U.S. Census Bureau, Census 2010.

[1] Design strata (44) are defined in Table 1-1.

7.3.3 American Community Survey for California

The 2014 and 2015 California American Community Survey public-use one-year micro data files (PUMS) were accessed for Dimensions 9-11. These data were used to estimate the proportions of the population by education, household tenure, and number of adults in the household within the seven California regions (Table 7-2). The 2014 and 2015 estimates were used for CHIS 2015 and CHIS 2016 one-year weights, respectively. The 2015 estimates were also used for the CHIS 2015-2016 two-year weights.

7.3.4 The National Health Interview Survey

The National Health Interview Survey (NHIS) is a primary source for estimates on household telephone service status (landline only, cell phone only, or dual user) for the U.S. as a whole and by state. Estimates for the state of California were obtained from the NHIS Early Release Program for years 2014

and 2015 to estimate telephone service type for CHIS 2015 and CHIS 2016, respectively. The estimates were required for calibration Dimension 12 and for combining landline and cell phone samples (see, for example, Section 4.5 related to adult weight adjustments) and are shown in Table 7-6.

Table 7-6. NHIS proportions of telephone service by person type and year

Person type	Telephone service[1]	Percentage	
		CHIS 2015[2]	CHIS 2016[3]
Adult	Landline only	7.2	5.6
	Dual use	49.0	46.4
	Cell phone only	43.8	48.0
Child and Teen	Landline only	4.0	3.2
	Dual use	44.7	41.0
	Cell phone only	51.3	55.8

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Dual use refers to households with both a landline and cellular telephone.

[2] Obtained from the National Health Interview Survey Early Release Program, 2014.

[3] Obtained from the National Health Interview Survey Early Release Program, 2015. Estimates also used for 2015-2016 CHIS two-year weights.

7.4 Producing the Control Totals

As mentioned previously, the population control totals were estimated and not directly drawn from available sources. The procedures to calculate the estimates follow methods develop for previous rounds of the study and are detailed below. The process begins with estimating and then removing population estimates linked with those living in group quarters (Section 7.4.1) and completes with the final calculations for the 13 calibration dimensions (Section 7.4.2).

7.4.1 Removing the Population Living in Group Quarters

Population control totals were not available and instead were estimated from the source information described previously. The procedures followed those originally developed for CHIS 2003 to maintain consistency across years. All control totals were derived from the same adjusted DOF projections to maintain consistency across dimensions. The general steps are described below.

Tabulated Population Projections. The DOF population counts were tabulated into groups defined by the cross-tabulation design stratum (44), ethnicity (Latino, Non-Latino), age group (18), race (6) and gender (2). The seven levels for race is shown in Table 7-4. The 18 age levels were required for

the calibration dimensions and are shown in Table 7-7. For convenience, let T_{d6}^{DOF} represent the cross-tabulated counts for the DOF file, where year is suppressed for convenience and the race grouping (6) excluding "other".

Estimated Group Quarters. The estimated proportion of group quarters was estimated from the 2010 Census SF1. As shown in Table 7-5, however, not all characteristics required for CHIS were available (e.g., single year of age). Consequently, assumptions were required: 1) the proportion in group quarters by single year of age within each age group (less than 18 years old, 18 to 64 years old, and 65 years old or older) was the same; and 2) the proportion in group quarters within racial group was the same across ethnicity (Latino or non-Latino).

Three sets of estimated control totals excluding group quarters were calculated from the 2010 Census SF1 by different groups. The first control total set was defined as

$$D_{1m}^{SF1.\overline{GQ}} = D_{1m}^{SF1} - D_{1m}^{SF1.GQ} \qquad (7.1)$$

where D_{1m}^{SF1} was the total population in California within group *m*, $D_{1m}^{SF1.GQ}$ was the corresponding population living in group quarters, and *m* was defined as cells created by crossing strata (44), race (7), age group (3), and sex (2). The levels for these variables are shown in Table 7-5.

Table 7-7 Age levels used to summarize California DOF data files

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 years and older

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Note: DOF = Department of Finance.

The second set of controls were defined as

$$D_{2p}^{SF1.\overline{GQ}} = D_{2p}^{SF1} - D_{2p}^{SF1.GQ} \qquad (7.2)$$

where D_{2p}^{SF1} was the total population in California within group *p*, $D_{2p}^{SF1.GQ}$ was the corresponding population living in group quarters, and *p* was defined as cells created by the cross of strata (44), ethnicity (3), age group (3), and sex (2).

The third set of controls were calculated as

$$D_{3q}^{SF1.\overline{GQ}} = D_{3q}^{SF1} - D_{3q}^{SF1.GQ} \tag{7.3}$$

where D_{3q}^{SF1} was the total population in California within group *q*, $D_{3q}^{SF1.GQ}$ was the corresponding population living in group quarters, and *q* was defined as cells created by the cross of strata (44) and age group (less than 18 years old, 18 years and older).

Using the similarity assumptions above and the three sets of control totals— $D_{1m}^{SF1.\overline{GQ}}$ in (7.1), $D_{2p}^{SF1.\overline{GQ}}$ in (7.2) and $D_{3q}^{SF1.\overline{GQ}}$ in (7.3)—that all excluded group quarters, 2010 Census SF1 counts with group quarters removed were estimated as

$$T_{d7}^{SF1.\overline{GQ}} = T_{mp}^{SF1} \times a_{mp} \tag{7.4}$$

where T_{mp}^{SF1} were the 2010 Census SFI population counts within cross-classified groups defined in Table 7-5, a_{mp} was the adjustment applied based on raking the counts to the control totals, and *d+* identifies the groups defined by the cross-classification of design stratum (44), ethnicity (Latino, Non-Latino), age group (18), race (7) including "other" and gender (2). The corresponding methodology was applied with the total population counts including group quarters to derive T_{d7}^{SF1}. Thus, the proportion of group quarters in cell *d* was calculated as

$$p_{d7}^{SF1.\overline{GQ}} = T_{d7}^{SF1.\overline{GQ}} \Big/ T_{d7}^{SF1} \tag{7.4}$$

This proportion was then applied to the yearly DOF files where ratios associated with the "other" category were assumed to be equivalent to a combination of information from the other racial groups (see, for example, *CHIS 2013-2014 Methodology Series: Report 5 – Weighting and Variance Estimation* for the justification). Thus,

$$T_{d6}^{DOF.\overline{GQ}} = p_{d7}^{SF1.\overline{GQ}} \times T_{d6}^{DOF} \tag{7.5}$$

the estimated residential population, excluding group quarters, within cells defined by stratum (44), ethnicity (Latino, Non-Latino), age group (18), race (6) and gender (2). The estimated proportion of the California residential population that live in grouped quarters was 2.5%.

7.4.2 Computing the Control Totals

Values calculated with (7.5) were tabulated across the estimation cells to form the non-group quarters control totals for calibration dimensions 1-3, 5, 6, and 8. Census tract information was used to align the 2010 Census SF1 file to SPA and San Diego HSR to form subarea-specific proportions. These were applied to the Los Angeles and San Diego adjusted counts for tabulate control totals for Dimension 4. For Dimension 7, the proportion by ethnicity group (Latino, non-Latino) for the Asian population was tabulated from 2010 Census SF2 file and applied to the adjusted DOF counts. ACS estimate proportions were used for Dimensions 9 (adult's education), 10 (number of adults in the household), 11 (non-telephone adjustment), and 13 (Imperial County adjustment). NHIS estimated proportions, used for combining landline and cell phone respondents, were again borrowed for form control totals for Dimension 12 (telephone usage).

8. IMPUTATION PROCEDURES

Item nonresponse occurs when a sample member should have but does not provide a response to a question. This excludes items that are skipped because of responses to prior routing questions. Item nonresponse also results if a response is deemed infeasible based on quality reviews and removed. Imputation replaces the missing values with valid responses, thereby enabling complete-case analysis and analysis weight creation. Imputation procedures were used for a select set of variables for CHIS 2015-2016.

This chapter describes the magnitude of item nonresponse by year for variables critical to producing the CHIS analysis weights, along with methods to address the missing information. Section 8.1 contains a preview of the variables subject to imputation, along with details of the methods used to supply the missing information. Identification of the methods used is communicated to the user community through a set of imputation indicator variables accompanying the data. Section 8.2 summarizes the imputation results for variables associated with the geographic location of the sampled households. Information on imputed values for household characteristics relevant to all interviews within the household (adult, teen, and child) is given in Section 8.3. Section 8.4 concludes this chapter with a discussion of the person-level variables important not only for the weights but also subgroup estimation with the CHIS data.

8.1 Imputed Variables and Methods

Table 8-1 lists by type the variables critical to the creation of CHIS analysis weights that were examined for imputation. The questionnaire response variables used to generate the initial values are provided. The response variables are listed in priority order, where priority was based on response source. For example, we assigned self-reported age (SRAGE) for adults the value from adult interview (AAGE); if this information was missing, then information was obtained from the child-first variable (KAGE) followed by the corresponding screener variable (AGE_R1-AGE_R20).

Table 8-1 also shows the order in which groups of weighting variables were imputed. Variables with the same group number were imputed simultaneously to ensure consistency of the resulting response values.

The sequence in which the variables were imputed was determined by the item nonresponse rate and association between the weighting variables. An item nonresponse rate was calculated as the number of valid responses for the variable divided by the total number of valid responses that should have been

recorded. If variable was needed for imputation of another because of a known or determined association, then that variable was imputed earlier in the process. Because of this, variables were not imputed by order of the "variable type" presented in the table. Otherwise, variables with smaller item nonresponse rates were imputed before those with larger rates. For this reason, the group order differs by year of CHIS.

Table 8-1. Description of imputed variables by year

Variable Type	Variable Name	Variable Description	Response Variables	Group Order by Year	
				2015	2016
Geographic (reported)	SR_COUNTY_FIPS	County	AH42, KAH42, SAH42	1	1
	SRZIP	ZIP Code	AM7, KAM7, SAM7	1	1
	SRSTRATA	Stratum	AH42, KAH42, SAH42	2	1
	SR_LASPA	Los Angeles Service Planning Area (SPA)	AH42, KAH42, SAH42, AM7, KAM7, SAM7	3	1
	SR_HR	San Diego Health Service Region (HSR)	AH42, KAH42, SAH42, AM7, KAM7, SAM7	3	1
Household	SRTENR	Household tenure	AK25, KAK25	12	9
	HASCELL	Cell/Wireless telephone service	AM33, KAM33	13	12
	HASLANDLINE	Landline telephone service	AN6, AN7	13	12
	CALLINTENSITY	Phone Use Intensity	AM34, HASCELL, HASLANDLINE	14	13
	POVERTY	Household Poverty Level	POVRT50, POVRT100, POVRT133, POVRT200, POVRT300, POVRT400	6	4
	ELIG_KID_0_5	Number of interview-eligible kids ages 0-5	GAR_1-GAR_20, GSP_1-GSP_20, AGE_1-AGE_20, GAGE_1-GAGE_20, SC14C_R1-SC14C_R20, SC14A_R1-SC14A_R20, SC14B_R1-SC14B_R20, ADULT_INDEX, TEEN_INDEX, CHILD_INDEX	15	10

(continued)

Table 8-1. Description of imputed variables by year (continued)

Variable Type	Weighting Variable Name	Weighting Variable Description	Response Variables	Group Order by Year	
				2015	2016
	ELIG_KID_6_11	Number of interview-eligible kids ages 6-11	GAR_1-GAR_20, GSP_1-GSP_20, AGE_1-AGE_20, GAGE_1-GAGE_20, SC14C_R1-SC14C_R20, SC14A_R1-SC14A_R20, SC14B_R1-SC14B_R20, ADULT_INDEX, TEEN_INDEX, CHILD_INDEX	15	10
	ELIG_TEEN	Number of interview-eligible teens	GAR_1-GAR_20, GSP_1-GSP_20, AGE_1-AGE_20, GAGE_1-GAGE_20, SC14C_R1-SC14C_R20, SC14A_R1-SC14A_R20, SC14B_R1-SC14B_R20, ADULT_INDEX, TEEN_INDEX, CHILD_INDEX	15	10
	PARENT_CHILD_HH	Number of parents for the selected child	GAR_1-GAR_20, GSP_1-GSP_20, AGE_1-AGE_20, GAGE_1-GAGE_20, SC14C_R1-SC14C_R20, SC14A_R1-SC14A_R20, SC14B_R1-SC14B_R20, ADULT_INDEX, TEEN_INDEX, CHILD_INDEX	15	11
	PARENT_TEEN_HH	Number of parents for the selected teen	GAR_1-GAR_20, GSP_1-GSP_20, AGE_1-AGE_20, GAGE_1-GAGE_20, SC14C_R1-SC14C_R20, SC14A_R1-SC14A_R20, SC14B_R1-SC14B_R20, ADULT_INDEX, TEEN_INDEX, CHILD_INDEX	15	11

(continued)

Table 8-1. Description of imputed variables by year (continued)

Variable Type	Weighting Variable Name	Weighting Variable Description	Response Variables	Group Order by Year	
				2015	2016
Person	SRAGE	Age	AAGE, AA2A, KAA2, KAA2A, AGE_R1-AGE_R20, CA3, TA2, CH_AGE_R1-CH_AGE_R20	5	3
	SRSEX	Sex	AA3, CA1, TA3, KAA3	4	2
	SREDUC	Educational Attainment	AH47, KAK47	11	8
	SRH	Self-Reported Latino	AA4, CH1, TI1	6	4
	SRW	Self-Reported White	AA5A_6, CH3_6, TI2_6	6	4
	SRAA	Self-Reported African American	AA5A_5, CH3_5, TI2_5	6	4
	SRAS	Self-Reported Asian	AA5A_4, CH3_4, TI2_4	6	4
	SRAI	Self-Reported American Indian/Alaska Native	AA5A_3, CH3_3, TI2_3	6	4
	SRPI	Self-Reported Native Hawaiian and Other Pacific Islander	AA5A_1, AA5A_2, CH3_1, CH3_2, TI2_1, TI2_2	6	4
	SRO	Self-Reported Other	AA5A_7, CH3_7, TI2_7	6	4
	SRCH	Self-Reported Chinese	SRAS2, AA5E_1-AA5E_18, TI7_1-TI7_18, CH7_1-CH7_18	9	6
	SRPH	Self-Reported Filipino	SRAS2, AA5E_1-AA5E_18, TI7_1-TI7_18, CH7_1-CH7_18	9	6
	SRKR	Self-Reported Korean	SRAS2, AA5E_1-AA5E_18, TI7_1-TI7_18, CH7_1-CH7_18	9	6
	SRJP	Self-Reported Japanese	SRAS2, AA5E_1-AA5E_18, TI7_1-TI7_18, CH7_1-CH7_18	9	6
	SRVT	Self-Reported Vietnamese	SRAS2, AA5E_1-AA5E_18, TI7_1-TI7_18, CH7_1-CH7_18	9	6
	SRASO	Self-Reported Other Asian	SRAS, AA5E_1-AA5E_18, TI7_1-TI7_18, CH7_1-H7_18	9	6

(continued)

Table 8-1. Description of imputed variables by year (continued)

Variable Type	Weighting Variable Name	Weighting Variable Description	Response Variables	Group Order by Year	
				2015	2016
	OMBSRREO	OMB Race/ Ethnicity Group	AA5A_A-AA5A_G, SRH, SRO, SRW2, SRAA2, SRAS2, SRAI2, SRPI2	8	5
	OMBSRASO	OMB non-Latino Asian Group	SRH, SRAS2, SRCH, SRPH, SRKR, SRJP, SRVT, SRASO	10	7

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

The type and item nonresponse rate of each variable dictated the imputation methodology. The various methods used for CHIS are shown in Table 8-2, along with the codes for the imputation indicator (flag) created for each weighting variable.

Table 8-2. Description of imputation indicators

Imputation Flag	Definition
0	Reported data; no imputation
1	Missing data; deterministic (i.e., logical) imputation[1]
2	Inconsistent data removed; deterministic (i.e., logical) imputation[1]
3	Missing data; random assignment[2]
4	Inconsistent data; random assignment[2]
5	Missing data; hot-deck imputation[3]
6	Inconsistent data; hot-deck imputation[3]
7	Missing data; external data source assignment
8	Inconsistent data; external data source assignment

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Values assigned based on other information in the interview

[2] Values randomly assigned from distribution of all possible values

[3] Values randomly obtained from donor record with reported data

A brief description of the imputation methods is as follows.

- *Deterministic imputation* uses responses to other variables within the respondent interview to assign a value. An example of deterministic imputation is imputing a female gender when the respondent has indicated a past pregnancy.

- *Random assignment* consists of randomly populating a value in place of the missing information based on the distribution of responses for that variable. One example of a random assignment is imputing a missing age based on the distribution of respondent ages in a stratum. Only variables with very few missing responses were imputed using deterministic or random assignment. While the item nonresponse may be related to other variables in the dataset, we assumed that any bias introduced through deterministic or random assignment would be negligible.
- *Hot-deck imputation* was used when the concerns about estimated bias from item nonresponse outweighed the applicability of the two imputation methods previously discussed. In hot-deck imputation, records with missing values are given values from randomly selected donors that were in the same imputation class as the recipient (RTI 2012; Andridge and Little, 2010; Brick and Kalton, 1996). Imputation classes are ideally formed through the cross-classification of covariates (variables) associated with the weighting variables in the group and with patterns of item nonresponse. We used results from classification and regression tree (CART) models to create imputation classes (Breiman et al., 1984) with input variables shown in Table 8-3.
- We imputed missing values using a *data source external to CHIS*, including population patterns derived from administrative data.

Table 8-3. Input variables for CART models to create imputation classes

Variable	Definition
ADLTFLG	Number of adults in the household
CALLINTENSITY	Self-reported phone intensity
CHLDFLG	Presence of children in the household
CREGION	California region
ELIG_KID_0_5	Number of children aged 0-5 years related to the selected adult
ELIG_KID_6_11	Number of children aged 6-11 years related to the selected adult
ELIG_TEEN	Number of teens aged 12-17 years related to the selected adult
HASCELL	Presence of a cell phone belonging to the respondent
HASLANDLINE	Presence of a landline phone in the household
PARENT_CHILD_HH	Number of parents associated with the selected child
PARENT_TEEN_HH	Number of parents associated with the selected teen
POVERTY	Poverty status

(continued)

Table 8-3. Input variables for CART models to create imputation classes (continued)

Variable	Definition
RENT	Household Rent Indicator
SRAGE	Self-reported age
SREDUC	Self-reported educational attainment
SRH	Self-reported Latino
SRRACE	Self-reported race
SRSEX	Self-reported sex
SRSTRATA	Self-reported stratum
SRTENR	Self-reported tenure
TEENFLG	Presence of teens in the household

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Several quality evaluations were conducted on the data before and after imputation. For example, data were subjected to an extensive cleaning process to ensure consistency of the responses within an interview (internal response consistency) and across interviews within a household (external response consistency) for the donor cases. Once completed, we examined the imputed response for internal and external consistency.

8.2 Geographic Characteristics

Records were geocoded to specific latitude and longitude coordinates based on the interview responses. Not all records, however, were accurately geocoded because of item nonresponse. This section describes the geographic responses imputed when missing to allow coordinate assignment by the geocoding process.

8.2.1 Self-reported County and ZIP Code

Self-reported county of residence (SR_COUNTY_FIPS) and ZIP code (SRZIP) were calculated from geocoded information. Missing responses occurred when such information could not be assigned from the geocodes. The missing values were imputed using a hot-deck procedure with imputation covariates area code, design stratum and reported stratum. Table 8-4 shows the unweighted item nonresponse rates for these variables.

Table 8-4. Item nonresponse for self-reported county of residence and zip code by interview mode

Variable and Source of Data	All Modes		Interview Mode[1]			
			Cell		Landline	
	n	pct[2]	n	pct[2]	n	pct[2]
SR_COUNTY_FIPS (County of residence FIPS)						
Reported values	41,418	98.4	19,993	97.7	21,425	99.1
Imputed values	671	1.6	480	2.3	191	0.9
Total	42,089	100.0	20,473	100.0	21,616	100.0
SRZIP (Self-reported ZIP code)						
Reported values	41,418	98.4	19,993	97.7	21,425	99.1
Imputed values	671	1.6	480	2.3	191	0.9
Total	42,089	100.0	20,473	100.0	21,616	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of cases within interview mode and variable.

8.2.2 Self-reported Stratum and Substratum

As with SR_COUNTY_FIPS and SRZIP, self-reported stratum (SRSTRATA), self-reported Los Angeles Service Planning Areas (SR_LASPA) and self-reported San Diego Health Service Regions (SR_HR) were computed from geocodes assigned for the respondent records. Missing values occurred when geocodes were insufficient for assignment, and were imputed using the same procedure as SR_COUNTY_FIPS and SRZIP discussed previously. Table 8-5 shows the item nonresponse for these variables by interview mode.

Table 8-5. Item nonresponse for self-reported stratum, Los Angeles SPA, and San Diego HSR by interview mode

Variable and Source of Data[1]	All Modes		Interview Mode[1]			
			Cell		Landline	
	n	pct[2]	n	pct[2]	n	pct[2]
SRSTRATA (Self-reported stratum)						
Reported values	39,051	92.8	17,729	86.6	21,322	98.6
Imputed values	3,038	7.2	2,744	13.4	294	1.4
Total	42,089	100.0	20,473	100.0	21,616	100.0

(continued)

Table 8-5. Item nonresponse for self-reported stratum, Los Angeles SPA, and San Diego HSR by interview mode (continued)

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data[1]	n	pct[2]	n	pct[2]	n	pct[2]
SR_LASPA (Self-reported Los Angeles county service planning area)						
Reported values	41,418	98.4	19,993	97.7	21,425	99.1
Imputed values	671	1.6	480	2.3	191	0.9
Total	42,089	100.0	20,473	100.0	21,616	100.0
SR_HR (Self-reported San Diego county health service region)						
Reported values	41,418	98.4	19,993	97.7	21,425	99.1
Imputed values	671	1.6	480	2.3	191	0.9
Total	42,089	100.0	20,473	100.0	21,616	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of cases within interview mode and variable.

8.2.3 Self-reported Region and Urbanicity

Three additional geographic variables were created based on the results of the geographic imputation. CREGION groups counties into seven distinct regions (Table 7-2). URBAN is a variable that classifies all records in strata 1-15 as urban (URBAN=1) and the remaining records as rural (URBAN=2). URBAN_NHIS is the 2013 National Health Information Survey urban classification code set for California (see Appendix A for details). All three variables were created after the imputation of and are based on SRZIP; therefore, their imputation counts match those of SRZIP.

8.3 Household Characteristics

To calculate the household weights, the foundation for the person-level analysis weight, all participating households must have data for certain characteristics. Furthermore, the dual-frame design of CHIS requires that records in the frame overlap (i.e., dual landline and cell phone users) be identified prior to weighting. This section outlines the imputation methodology for these household variables.

8.3.1 Household Tenure

Missing values for household tenure (SRTENR) were imputed using hot-deck imputation. CART created imputation classes using household poverty (POVERTY) and self-reported Hispanic ethnicity (SRH), in addition to self-reported stratum. Table 8-6 shows the item nonresponse distribution for this variable by interview mode.

Table 8-6. Item nonresponse for self-reported household tenure by interview mode

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
SRTENR (Household tenure)						
Reported values	41,549	98.7	20,185	98.6	21,364	98.8
Imputed values	540	1.3	288	1.4	252	1.2
Total	42,089	100.0	20,473	100.0	21,616	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of cases within interview mode and variable.

8.3.2 Telephone Service

HASCELL indicates the presence of a cell phone belonging to the respondent regardless of the interview mode, while HASLANDLINE indicates the presence of a landline phone associated with the household. CALLINTENSITY classifies the average amount of use for each device. Missing values for these items were imputed using hot-deck imputation. Imputation classes for HASCELL and HASLANDLINE were created using CART from SRAGE and POVERTY, while the imputation classes for CALLINTENSITY were created from SRAGE and SRTENR. Furthermore, some records that reported only having a landline phone had also inconsistently reported making calls on a cell phone. These records had their CALLINTENSITY code logically imputed to be consistent with their reported landline phone presence. The item nonresponse for these variables is shown in Table 8-7.

Table 8-7. Item nonresponse for presence of cell phone, presence of landline phone, and type of phone usage by interview mode

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
HASCELL (Presence of a cell phone)						
Reported values	41,868	99.5	20,473	100.0	21,395	99.0
Imputed values	221	0.5	0	0.0	221	1.0
Total	42,089	100.0	20,473	100.0	21,616	100.0
HASLANDLINE (Presence of a landline phone)						
Reported values	41,626	98.9	20,013	97.8	21,613	100.0
Imputed values	463	1.1	460	2.2	3	0.0
Total	42,089	100.0	20,473	100.0	21,616	100.0
CALLINTENSITY (Self-reported phone intensity)						
Reported values	36,437	86.6	17,166	83.8	19,271	89.2
Imputed values	5,652	13.4	3,307	16.2	2,345	10.8
Total	42,089	100.0	20,473	100.0	21,616	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of cases within interview mode and variable.

8.3.3 Household Composition

Number of Eligible Children by Age Group

The number of children related to the adult respondent was required for household and child-level weights. Because children in different age groups had different probabilities of selection, we separated the number of eligible children by age group. Missing values were imputed using hot-deck imputation with reported stratum, the type of respondents (adult, child, or teen) in each household and the parent's race/ethnicity as imputation covariates. The item nonresponse for the two age-group variables is shown in Table 8-8.

Table 8-8. Item nonresponse for number of study-eligible children by age group and interview mode

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
ELIG_KID_0_5 (Self-reported number of eligible children age 0-5)						
Reported values	42,062	99.9	20,454	99.9	21,608	100.0
Imputed values	27	0.1	19	0.1	8	0.0
Total	42,089	100.0	20,473	100.0	21,616	100.0
ELIG_KID_6_11 (Self-reported number of eligible children age 6-11)						
Reported values	42,069	100.0	20,461	99.9	21,608	100.0
Imputed values	20	0.0	12	0.1	8	0.0
Total	42,089	100.0	20,473	100.0	21,616	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of cases within interview mode and variable.

Number of Eligible Teens

The number of teens related to the adult respondent was required for the household and teen-level weights. As with the 'number of eligible children by age group' methodology, hot-deck imputation was implemented using reported design stratum, type of respondents (adult, child, and teen) in the household, the race/ ethnicity of the selected parent, and a screener indicator on the presence of a child or teen in the household as imputation classes. The corresponding item nonresponse is shown in Table 8-9.

Table 8-9. Item nonresponse for number of study-eligible teens by interview mode

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
ELIG_TEEN (Self-reported number of eligible teens)						
Reported values	42,080	100.0	20,470	100.0	21,610	100.0
Imputed values	9	0.0	3	0.0	6	0.0
Total	42,089	100.0	20,473	100.0	21,616	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of cases within interview mode and variable.

Number of Parents of Selected Child or Teen

The number of parents in the household for the selected child and teen were used to construct the corresponding person-level weight. Prior to imputation, child-first interviews were logically coded to have two parents for the selected child and/or teen; the child-first methodology was implemented only when the screener parent differed from the other parent who was selected for the interview.

Missing values were imputed by hot-deck imputation with reported stratum (SRSTRATA), the family composition, and the parent's race/ethnicity used as imputation covariates. The item nonresponse for the two-parent variables is shown in Table 8-10.

Table 8-10. Item nonresponse for number of parents in household for child and teen by interview mode

	Interview Mode[1]					
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
PARENT_CHILD_HH (Self-reported number of selected child's parents)						
Reported values	1,946	45.3	1,429	55.4	491	29.2
Imputed values	2,347	54.7	1,151	44.6	1,191	70.8
Total[3]	4,293	100.0	2,580	100.0	1,682	100.0
PARENT_TEEN_HH (Self-reported number of selected teen's parents)						
Reported values	784	49.2	452	54.7	322	42.6
Imputed values	810	50.8	375	45.3	433	57.4
Total[4]	1,594	100.0	827	100.0	755	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of cases within interview mode and variable.

[3] Universe = completed child interviews.

[4] Universe = completed teen interviews.

8.3.4 Poverty Status

Poverty status was used in the CART models to develop imputation classes for other variables. This variable was not used in the weighting process. The item nonresponse rates shown in Table 8-11 were due to missing information for the components used to calculate poverty. Data for adult respondents who answered "unknown" to the household income questions were left unchanged; these responses were

slightly more prevalent in the landline sample. Missing values for the remaining seven respondents were imputed via random assignment within reported stratum.

Table 8-11. Item nonresponse for poverty status by interview mode

	Interview Mode[1]					
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
POVERTY (Self-reported poverty status)						
Reported data; no imputation	34,594	82.2	17,284	84.4	17,310	80.1
Random assignment - missing data	7	0.0	4	0.0	3	0.0
Unknown response	7,488	17.8	3,185	15.6	4,303	19.9
Total	42,089	100.0	20,473	100.0	21,616	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of cases within interview mode and variable.

8.4 Person-level Characteristics

Person-level weights are used to calculate population estimates for CHIS. However, the person-level variables contained item nonresponse among those classified as study respondents (Table 8-12). This section describes the imputation procedures used for each variable needed for weighting and their item nonresponse rates.

Table 8-12. Respondents by person type and interview mode

	Respondents by Interview Mode[1]				
	All Modes	Cell		Landline	
Person Type	n	n	pct[2]	n	pct[2]
Adult	42,089	20,473	48.6	21,616	51.4
Child	4,293	2,580	60.1	1,713	39.9
Teen	1594	827	51.9	767	48.1

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of respondents by interview mode and person type.

8.4.1 Sex and Age

Self-reported sex (SRSEX) and self-reported age (SRAGE) were derived from a combination of screener and interview variables for each respondent. Table 8-13 shows the item nonresponse for SRSEX and SRAGE for each type of respondent. Because the nonresponse rates were low for SRSEX, missing values were imputed using random assignment from the distribution of responses within the associated reported stratum. SRAGE was imputed by hot-deck imputation using stratum and screener age group classification as imputation classes.

Table 8-13. Item nonresponse for self-reported sex and age by person type and interview mode

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
SRSEX (Self-reported sex)						
Adult	8	0.0	8	0.0	0	0.0
Child	13	0.3	7	0.3	6	0.4
Teen	1	0.1	0	0.0	1	0.1
SRAGE (Self-reported age)						
Adult	269	0.6	77	0.4	192	0.9
Child	21	0.5	6	0.2	15	0.9
Teen	14	0.9	7	0.8	7	0.9

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of imputed records among respondents in Table 8-12 by mode and person type.

8.4.2 Race and Ethnicity

Single Race and Ethnicity

The seven self-reported race and ethnicity variables were created after upcoding all responses to the associated questions. Missing values for all variables were imputed by a single hot-deck imputation, using stratum as the imputation class. Because of the presence of multiple-race households, missing race and ethnicity values for child and teen respondents used the adult respondent's race/ethnicity combination as another imputation class. Table 8-14 shows the response patterns by interview mode and variable grouping for respondents missing at least one self-reported race or ethnicity value. Table 8-15 shows the response patterns for the self-reported race variables by interview mode.

Table 8-14. Item nonresponse for any self-reported race value and ethnicity by interview mode

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
One or more imputed Race values						
Adult	1,284	3.1	826	4.0	458	2.1
Child	265	6.2	163	6.3	102	6.0
Teen	135	8.5	83	10.0	52	6.8
SRH (Self-reported Latin ethnicity)						
Adult	195	0.5	79	0.4	116	0.5
Child	265	6.2	163	6.3	102	6.0
Teen	135	8.5	83	10.0	52	6.8

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of imputed records among respondents in Table 8-12 by mode and person type.

Table 8-15. Item nonresponse for single-response self-reported race by person type and interview mode

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
SRW (Self-reported race: White)						
Adult	1,238	2.9	807	3.9	431	2.0
Child	265	6.2	163	6.3	102	6.0
Teen	135	8.5	83	10.0	52	6.8
SRAA (Self-reported race: African American)						
Adult	1,251	3.0	812	4.0	439	2.0
Child	265	6.2	163	6.3	102	6.0
Teen	135	8.5	83	10.0	52	6.8
SRAI (Self-reported race: American Indian)						
Adult	1,250	3.0	811	4.0	439	2.0
Child	265	6.2	163	6.3	102	6.0
Teen	135	8.5	83	10.0	52	6.8

(continued)

Table 8-15. Item nonresponse for single-response self-reported race by person type and interview mode (continued)

Variable and Source of Data	All Modes		Interview Mode[1]			
			Cell		Landline	
	n	pct[2]	n	pct[2]	n	pct[2]
SRAS (Self-reported race: Asian)						
Adult	1,245	3.0	809	4.0	436	2.0
Child	265	6.2	163	6.3	102	6.0
Teen	135	8.5	83	10.0	52	6.8
SRPI (Self-reported race: Pacific Islander)						
Adult	1,250	3.0	812	4.0	438	2.0
Child	265	6.2	163	6.3	102	6.0
Teen	135	8.5	83	10.0	52	6.8
SRO (Self-reported race: Other)						
Adult	1,245	3.0	809	4.0	436	2.0
Child	265	6.2	163	6.3	102	6.0
Teen	135	8.5	83	10.0	52	6.8

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Interview mode is the type of telephone used during recruitment regardless of sampling frame type (e.g., ported landline telephones are classified as cell for mode of interview).

[2] Unweighted percent of imputed records among respondents in Table 8-12 by mode and person type.

OMB Race/Ethnicity Variable

The weighting algorithm calibrated the survey weights to match the California Department of Finance (DOF) population estimates for race and ethnicity. Since the DOF race and ethnicity estimates were based on the revised OMB 1997 standards for data collection, only five race categories are available: White, African American, Asian American Indian, and Pacific Islander. The 2010 Census race estimates included an additional category called "Other Race" for respondents who did not report their races in one of the five categories. To match the Office of Management and Budget (OMB) standards, the U.S. Census Bureau created a Modified Race Data Summary file (MRDSF) that recodes the "Other" respondents into one of the five OMB race codes. CHIS collected race data for the six Census race categories; therefore, the "Other" respondents need to be recoded into the five race categories. These race categories are coded into the variable OMBSRREO.

Table 8-16 shows the race classification for OMBSRREO. There are also classifications for respondents who identify as Latino and respondents who identify as belonging to multiple races. These last two classifications were included to reduce the number of records that require imputation for OMBSRREO.

Table 8-16. Classification codes for OMB self-reported race/ethnicity

OMBSRREO Code	Description
1	Latino
2	Non-Latino White Only
3	Non-Latino African American Only
4	Non-Latino American Indian Alaskan Native Only
5	Non-Latino Asian Only
6	Non-Latino Pacific Islander Native Hawaiian Only
7	Non-Latino Two or More Races

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

The same coding and imputation procedure consistent with prior years of CHIS was used to classify all records into the five OMB race categories. The imputed self-reported race and ethnicity variables (SRH, SRW, SRAA, SRAS, SRAI, SRPI, and SRO) were used for the coding process.

Another indicator variable, MULTIRACE, was created to identify records that reported two or more races. All respondents who self-identified as Latino (SHR = 1) were coded as such regardless of any other race indications. Non-Latino respondents who either self-identified as one of the OMB race categories or "Other" (SRO = 1), and one of the OMB race categories were assigned to that race category. Non-Latino respondents who reported two or more races (MULTIRACE = 1) or who only reported multiple instances of "Other" were classified as having two or more races. Non-Latino respondents who only reported "Other" were required to have an imputed OMB race.

The hot-deck imputation procedure required temporary race variables (SRW2, SRAA2, SRAI2, SRAS2, and SRPI2) created from the self-reported single race variables. Non-Latino respondents who only reported "Other" had these variables set as missing. No other types of records were marked to be imputed. Hot-deck imputation proceeded on these variables. Adult records used reported stratum, SRH, and self-reported Asian ethnicity (SRAS2) as imputation classes. Child and teen records used California region (CREGION) instead of reported stratum because of small sample sizes and parent's race/ethnicity. If a parent race/ethnicity category did not have enough records to provide donors for all regions (n=100 minimum), CREGION was dropped as an imputation class to preserve the parent-child race relationship.

Records were then classified into the OMB races based on the imputed data. Table 8-17 shows the results of the hot-deck procedure by interview mode, person type and OMBSRREO value.

Table 8-17. Item nonresponse for office and management and budget self-reported race/ethnicity by person type and interview mode

			Interview Mode			
	All Modes		Cell		Landline	
OMBSRREO Value, Person Type	n	pct[1]	n	pct[1]	n	pct[1]
Non-Latino White Only						
Adult	45	0.1	25	0.1	20	0.1
Child	2	0.0	1	0.0	1	0.1
Teen	0	0.0	0	0.0	0	0.0
Non-Latino African American Only						
Adult	4	0.0	3	0.0	1	0.0
Child	1	0.0	0	0.0	1	0.1
Teen	0	0.0	0	0.0	0	0.0
Non-Latino American Indian Alaskan Native Only						
Adult	2	0.0	2	0.0	0	0.0
Child	0	0.0	0	0.0	0	0.0
Teen	0	0.0	0	0.0	0	0.0
Non-Latino Asian Only						
Adult	5	0.0	1	0.0	4	0.0
Child	0	0.0	0	0.0	0	0.0
Teen	0	0.0	0	0.0	0	0.0
Non-Latino Pacific Islander Native Hawaiian Only						
Adult	1	0.0	1	0.0	0	0.0
Child	0	0.0	0	0.0	0	0.0
Teen	0	0.0	0	0.0	0	0.0
Non-Latino Two or More Races						
Adult	34	0.1	16	0.1	18	0.1
Child	5	0.1	2	0.1	3	0.2
Teen	0	0.0	0	0.0	0	0.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Unweighted percent of imputed records among respondents in Table 8-12 by mode and person type.

OMB Asian Ethnicity Group

Records identified as Asian by the temporary variable SRAS2 were then further classified by Asian ethnicity in the variable OMBSRASO. The seven classes in OMBSRASO are listed in Table 8-18.

Table 8-18. Classification codes for office and management and budget self-reported non-Latino Asian ethnicity

OMBSRASO Code	Asian Ethnicity Indicator Variable	Description
-1	*N/A*	Latino or Non-Asian
1	SRCH	Chinese Only
2	SRKR	Korean Only
3	SRPH	Filipino Only
4	SRVT	Vietnamese Only
5	SRASO	Other Asian Ethnicity
6	SRJP	Japanese Only

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.
N/A = not applicable.

After imputation for SRAS2, six Asian ethnicity indicator variables were created based on their responses to the Asian ethnicity questions. Hot-deck imputation proceeded on these temporary variables. Adult records used reported stratum, SRH, and SRAS2 as imputation classes. Child and teen records used CREGION instead of stratum because of small sample sizes in certain strata, and parent race/ethnicity. If a parent race/ethnicity combination did not have enough records to provide donors for all regions (100 records was the cut-off level), CREGION was dropped as an imputation class to preserve the parent-child race relationship. Table 8-19 shows the results of the hot-deck procedure on the single-race Asian ethnicity variables by interview mode and person type.

Records were then coded into OMBSRASO based on their imputed Asian ethnicity variables. Table 8-20 shows the results of the hot-deck procedure by interview mode, person type and OMBSRASO value.

Table 8-19. Item nonresponse for single-response self-reported non-Latino Asian ethnicity by person type and interview mode

	All Modes		Interview Mode			
			Cell		Landline	
Single race, Person Type	n	pct[1]	n	pct[1]	n	pct[1]
SRCH (OMB Asian ethnicity: Chinese)						
Adult	94	0.2	44	0.2	50	0.2
Child	48	1.1	20	0.8	28	1.6
Teen	18	1.1	7	0.8	11	1.4
SRKR (OMB Asian ethnicity: Korean)						
Adult	97	0.2	45	0.2	52	0.2
Child	37	0.9	15	0.6	22	1.3
Teen	9	0.6	3	0.4	6	0.8
SRPH (OMB Asian ethnicity: Filipino)						
Adult	97	0.2	45	0.2	52	0.2
Child	43	1.0	17	0.7	26	1.5
Teen	13	0.8	5	0.6	8	1.0
SRVT (OMB Asian ethnicity: Vietnamese)						
Adult	97	0.2	45	0.2	52	0.2
Child	47	1.1	21	0.8	26	1.5
Teen	18	1.1	7	0.8	11	1.4
SRASO (OMB Asian ethnicity: Asian Other)						
Adult	97	0.2	45	0.2	52	0.2
Child	35	0.8	13	0.5	22	1.3
Teen	13	0.8	5	0.6	8	1.0
SRJP (OMB Asian ethnicity: Japanese)						
Adult	94	0.2	44	0.2	50	0.2
Child	34	0.8	14	0.5	20	1.2
Teen	9	0.6	3	0.4	6	0.8

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Unweighted percent of imputed records among respondents in Table 8-12 by mode and person type.

Table 8-20. Item nonresponse for office and management and budget self-reported non-Latino Asian ethnicity by person type and interview mode

			Interview Mode			
	All Modes		Cell		Landline	
OMBSRASO, Person Type	n	pct[1]	n	pct[1]	n	pct[1]
Chinese only						
Adult	35	0.1	21	0.1	14	0.1
Child	5	0.1	1	0.0	4	0.2
Teen	0	0.0	0	0.0	0	0.0
Korean only						
Adult	5	0.0	4	0.0	1	0.0
Child	2	0.0	0	0.0	2	0.1
Teen	0	0.0	0	0.0	0	0.0
Filipino only						
Adult	14	0.0	7	0.0	7	0.0
Child	13	0.3	5	0.2	8	0.5
Teen	14	0.9	7	0.8	7	0.9
Japanese only						
Adult	13	0.0	6	0.0	7	0.0
Child	5	0.1	2	0.1	3	0.2
Teen	0	0.0	0	0.0	0	0.0
Other Asian ethnicity						
Adult	15	0.0	4	0.0	11	0.1
Child	11	0.3	6	0.2	5	0.3
Teen	0	0.0	0	0.0	0	0.0
Vietnamese only						
Adult	14	0.0	3	0.0	11	0.1
Child	14	0.3	8	0.3	6	0.4
Teen	4	0.3	0	0.0	4	0.5

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Unweighted percent of imputed records among respondents in Table 8-12 by mode and person type.

8.4.3 Educational Attainment

Missing values for the educational attainment of the selected adult (SREDUC) were imputed using a hot-deck method (Table 8-21). A CART analysis identified the imputation covariates as POVERTY, SRH and reported stratum.

Table 8-21. Item nonresponse for self-reported educational attainment of the adult by person type and interview mode

			Interview Mode			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[1]	n	pct[1]	n	pct[1]
SREDUC (Self-reported educational attainment)						
Reported values	41,825	99.4	20,350	99.4	21,475	99.3
Imputed values	264	0.6	123	0.6	141	0.7
Total	42,089	100.0	20,473	100.0	21,616	100.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Unweighted percent of cases within interview mode and variable.

9. VARIANCE ESTIMATION

Weights detailed in chapters 4-6 are used to generate point estimates from CHIS data. In this chapter, we discuss the calculation of precision for those estimates, most notably quantified through a standard error or the square root of the sampling variance. Section 9.1 summarizes the precision for a select number of analysis variables from the adult, child, and teen analysis files. Section 9.2 discusses two types of variance estimation methods that may be used for CHIS—linearization and replication. We detail the creation of the values needed for replication variance estimation in Section 9.3. This chapter concludes in Section 9.4 with information relevant for calculating estimates with standard commercial and open-source software that properly account for the CHIS sampling design.

9.1 Design Effects

Point estimates are only part of the story for any survey. Measures of precision, most notably the sampling error, quantify the confidence one has that a point estimate is a good representation of the true (but unknown) population parameter. For example, estimates with a small standard error (and consequently relatively high precision) are viewed more variably than those with low precision because they enable tests of significance. Though point estimates appear to be substantively different, their large standard errors may result in an insignificant statistical test of those differences.

There are several statistics for quantifying precision of an estimate. They include:

- the standard error, or SE, defined as the square root of the sampling variance for an estimate that is specific to the survey design;
- the coefficient of variation, or CV, defined as the SE of the estimates divided by the point estimate;
- the relative variance, or relvariance, defined as squared CV;
- the confidence interval calculated as the range of values from the lower bound (the point estimate minus a specified multiple of SE) to the upper bound (the point estimate plus the specified multiple of SE used for the lower bound); and
- the design effect, described below.

The design effect (DEFF) was developed by Leslie Kish (1965). DEFF typically quantifies the increase in a SE for an estimate from a complex sample design above the SE calculated for a single stage stratified design (stsrs) with sample proportionally allocated to strata as distributed in the population. A

stsrs design is considered optimal for small SEs; deviations from this design are generally implemented to meet analytic objectives such as relatively equal sample across strata in CHIS.

DEFF for an estimate $\hat{\theta}$ is calculated as

$$\frac{\text{var}_{\pi}\left(\hat{\theta}\right)}{\text{var}_{stsrs}\left(\hat{\theta}\right)} \tag{9.1}$$

where $\text{var}_{\pi}\left(\hat{\theta}\right)$ is the variance estimate for the appropriate CHIS sample design, and $\text{var}_{stsrs}\left(\hat{\theta}\right)$ is the variance for the stsrs design. Variance for the CHIS sample design, $\text{var}_{\pi}\left(\hat{\theta}\right)$, accounts for the following aspects of the survey design using replication methods discussed in this chapter:

- **Design strata**. Mutually exclusive stratification variables for CHIS were county or county group within California drawn from multiple sampling frames (landline, cell, surname and ABS).
- **Clustering**. Analyses involving the combination of adult and child interviews would result in household-clustered estimates.
- **Over- and under-sampling of sample members**. Deviations from sampling proportional to the distribution in the population will result in either over- or under-sampling of subgroups in the population. The CHIS 2015-2016 targets included an equal allocation to landline and cell phone samples; however, certain strata had a higher proportion of cell-only households, resulting in an under-sample of those sample members. A higher proportion of persons with one or more Asian nationalities were recruited for CHIS for specialized analyses, thereby introducing over-sampling for this subgroup.
- **Household Subsampling**. New to CHIS 2015-2016 was subsampling of nonrespondents for nonresponse follow-up (NRFU; see Section 3.3.6 in *CHIS 2015-2016 Methodology Series: Report 1 – Sample Design*).
- **Within-Household Subsampling**. Subsampling within CHIS households occurred for those with multiple adult residents contacted through a randomly chosen landline telephone number, for households with multiple eligible children, and for households with multiple eligible teens. Child and teen subsampling occurred regardless of frame from which the telephone number was chosen.
- **Weight Adjustments**. As discussed in the previous sections of this report, differential weight adjustments were applied to reduce nonresponse bias and additional coverage bias not

addressed through the nonresponse adjustments. Additionally, composite adjustments were used to combine landline and cell phone samples (CHIS 2015-2016) and to combine ABS with landline/cell samples (CHIS 2016 only)

SUDAAN provides four DEFF calculations (RTI, 2012, Chp. 12). The DEFF formulation used for the statistics presented below uses $\text{var}_{stsrs}\left(\hat{\theta}\right)$ in equation (9.1) derived from a simple random sample with no analysis weights, no strata (or only one design stratum), and no clustering.

In days past, DEFF was used to adjust estimates from software that could only calculate SEs for a stsrs design. Specialized software for analyzing survey data obtained through a complex, multistage design is widely available now. Hence, DEFF is most effectively used to compare before and after a weight adjustment is applied (as implemented for CHIS 2015-2016) or across multiple rounds of a survey using the same sampling design. Thus, differences in DEFF between CHIS 2015-2016 and prior rounds of the study cannot be easily explained as changes to the sampling design, weighting methodology, differential response, and the like will result in different precision estimates.

As in past rounds, CHIS DEFFs calculated for specific variables of interest will generally have values greater than one. This is typical for surveys with complex designs and weighting schemes, and with over- and under-sampling to achieve analytic objectives. The degree of deviations from one will differ by the type of estimate. For example, characteristics that are linearly associated with the calibration controls used in the CHIS final weighting step will have lower DEFFs than those with weaker associations (see, e.g., Valliant et al., 2013).

Because precision differs by questionnaire item, tables below summarize DEFF for a series of variables from the adult, teen and child questionnaires. Specifically, the average, maximum and minimum DEFFs are shown by person interview overall and by reported stratum are shown. Because the distribution of DEFFs are known to be non-symmetric, the median values are also provided. Finally, the average square root of DEFF, denoted as DEFT, is listed along with the other measures. DEFT aligns with SE (instead of variance as with DEFF) and also provides some measure of smoothing if the DEFFs from the set of questionnaire items analyzed vary widely

Tables 9-1, 9-2, and 9-3 contain DEFFs and DEFTs for items selected from the adult, child and teen questionnaires, respectively. Each table contains the average, median, maximum and minimum DEFF along with the average DEFT, overall and by reported stratum. All calculations used the final person-level linear weights described in the previous chapters.

A total of 37 variables were chosen for the adult DEFF analyses (Table 9-1). The variables include health characteristics such as general health rating, diagnosis (asthma, diabetes, high blood pressure, heart failure/congestive, heart disease, difficulty learning and remembering, blind/deaf, level of physical impairment, felt nervous), lifestyle (smoking, alcohol use, had fast food, number of sexual partners, skipped meals, feel safe), preventive medicine (flu vaccine, delayed medical care, usual source of healthcare, number of doctor visits), health insurance (insured, Medicare/Medi-CAL, financial assistance, employer health insurance, other government health plan, prescription coverage), and socioeconomic and demographic variables (income, sexual orientation, marital status, education attainment, country of birth, U.S. citizenship status). The average DEFT for CHIS 2015 was 1.84 overall and ranged from 1.41 to 2.56 across the reported strata. These values are slightly higher for CHIS 2016—2.34 overall and 1.45 to 2.61 across reported stratum.

A total of 24 variables were chosen for the child DEFF analyses (Table 9-2). These variables include health characteristics such as general health rating, diagnosis (asthma, child visited emergency room), lifestyle (go to the park, park safety concerns, had fast food, access to the internet), preventive medicine (usual healthcare location, doctor visits, flu vaccine, delayed medical care/medication, access to childcare), financial assistance with health insurance, and socio economic and demographic variables (servings of juice, fruit and vegetables, parents' country of origin). The average DEFT for CHIS 2015 was 1.67 overall and ranged from 1.03 to 1.54 across the reported strata. These values are slightly higher for CHIS 2016—1.91 overall and 1.00 to 1.79 across reported stratum.

A total of 23 variables were chosen for the teen DEFF analyses (Table 9-3). These variables include health characteristics such as general health rating, diagnosis (asthma, teen visited emergency room, felt nervous, had psychological or emotional counseling, physical abuse), lifestyle (smoking, alcohol use, go to the park, park/school safety concerns, had fast food), preventive medicine (usual healthcare location, doctor visits, delayed medical care/medication, physical activity, sexually active), and socio economic and demographic variables (servings of vegetables, school attendance). The average DEFT for CHIS 2015 was 1.81 overall and ranged from 1.03 to 1.54 across the reported strata. These values are slightly higher for CHIS 2016—2.66 overall and 0.92 to 2.16 across reported stratum. Note that stratum estimates were suppressed because of small sample sizes leading to unstable estimates.

Table 9-1. Design effect (DEFF) and square root DEFF (DEFT) statistics for estimates from the adult interviews, overall and by reported stratum within study year

Reported stratum	CHIS 2015					CHIS 2016				
	DEFF				Average DEFT	DEFF				Average DEFT
	Average	Median	Max	Min		Average	Median	Max	Min	
State	3.47	3.01	6.73	1.96	1.84	5.63	4.90	13.09	3.01	2.34
1 – Los Angeles	2.90	2.33	7.14	1.36	1.67	4.44	3.75	13.20	1.98	2.06
2 – San Diego	2.56	2.17	8.61	1.70	1.57	4.10	3.45	14.01	1.08	1.94
3 – Orange	2.78	2.60	5.86	1.56	1.65	4.88	4.11	16.22	1.56	2.14
4 – Santa Clara	2.72	2.50	5.37	1.47	1.63	5.06	4.75	12.37	2.18	2.21
5 – San Bernardino	2.86	2.31	11.47	1.27	1.63	4.44	4.11	7.79	1.94	2.08
6 – Riverside	2.69	2.44	6.18	1.38	1.62	3.87	3.51	7.45	1.45	1.93
7 – Alameda	2.97	2.96	5.25	1.72	1.71	4.53	4.39	8.55	1.98	2.09
8 – Sacramento	3.47	3.01	13.64	1.47	1.81	4.15	4.10	7.83	1.76	2.02
9 – Contra Costa	2.47	2.18	6.62	1.53	1.55	3.08	3.06	5.60	0.97	1.73
10 – Fresno	2.60	2.63	4.29	1.23	1.60	3.89	3.62	7.26	1.28	1.94
11 – San Francisco	3.57	3.25	7.15	2.18	1.87	5.28	4.72	13.51	2.49	2.25
12 – Ventura	2.61	2.58	4.34	1.52	1.60	4.11	3.72	13.23	0.97	1.96
13 – San Mateo	2.35	2.38	3.32	1.38	1.52	4.14	4.12	10.90	0.78	1.97
14 – Kern	2.04	1.82	4.89	1.19	1.41	3.83	2.98	20.33	1.68	1.87
15 – San Joaquin	2.04	1.98	3.54	1.16	1.42	3.48	2.98	9.70	1.48	1.82
16 – Sonoma	2.87	2.84	6.06	0.92	1.67	3.90	3.64	10.13	0.71	1.90
17 – Stanislaus	2.28	2.11	5.02	1.01	1.49	2.84	2.69	6.60	1.70	1.67
18 – Santa Barbara	2.04	2.08	3.58	0.83	1.41	2.98	2.89	7.58	0.70	1.68
19 – Solano	2.42	2.39	3.94	1.06	1.54	3.36	3.29	6.68	1.22	1.80
20 – Tulare	2.71	2.45	6.18	0.83	1.61	2.86	2.76	6.02	1.12	1.67
21 – Santa Cruz	2.72	1.92	10.85	1.12	1.58	2.91	2.47	6.17	1.41	1.68
22 – Marin	7.38	5.76	22.54	1.76	2.56	3.43	2.63	9.94	0.70	1.77
23 – San Luis Obispo	2.51	2.61	5.14	1.01	1.55	3.66	3.29	7.80	1.13	1.87
24 – Placer	2.22	2.11	3.27	1.40	1.48	2.88	2.80	5.64	1.21	1.68

(continued)

9-5

Table 9-1. Design effect (DEFF) and square root DEFF (DEFT) statistics for estimates from the adult interviews, overall and by reported stratum within study year (continued)

Reported stratum	CHIS 2015					CHIS 2016				
	DEFF				Average DEFT	DEFF				Average DEFT
	Average	Median	Max	Min		Average	Median	Max	Min	
25 – Merced	2.70	2.47	6.11	0.87	1.62	4.55	4.12	16.93	0.85	2.02
26 – Butte	3.00	2.79	6.45	1.25	1.70	2.90	2.91	5.45	0.86	1.66
27 – Shasta	2.21	1.92	5.95	1.12	1.47	2.92	2.60	7.69	1.20	1.67
28 – Yolo	4.63	4.56	9.98	1.25	2.10	3.66	3.22	10.00	0.87	1.81
29 – El Dorado	3.35	2.88	10.41	1.15	1.76	7.41	6.82	15.81	1.33	2.61
30 – Imperial	2.54	1.98	8.31	1.00	1.54	4.04	3.68	9.81	1.47	1.96
31 – Napa	3.96	3.81	9.00	1.49	1.94	6.06	6.25	16.64	1.04	2.36
32 – Kings	2.97	3.09	7.82	0.76	1.69	4.25	4.11	11.37	1.06	1.99
33 – Madera	2.14	2.15	5.08	0.87	1.45	4.99	4.06	14.91	1.83	2.16
34 – Monterey	3.19	3.03	5.98	1.98	1.77	3.13	2.85	6.94	1.73	1.75
35 – Humboldt	2.41	2.28	6.09	1.06	1.53	3.33	2.98	17.96	0.90	1.73
36 – Nevada	2.76	2.80	5.16	1.13	1.64	2.13	2.07	4.16	0.76	1.45
37 – Mendocino	2.16	2.07	4.91	0.98	1.45	3.53	3.27	7.78	0.92	1.82
38 – Sutter	5.31	4.74	16.16	1.26	2.15	3.69	3.31	15.21	1.04	1.85
39 – Yuba	2.36	1.92	6.85	1.01	1.50	3.33	3.13	6.18	1.51	1.80
40 – Lake	2.79	2.55	8.27	0.80	1.63	3.13	3.17	9.54	1.31	1.73
41 – San Benito	3.70	3.12	9.85	0.94	1.84	4.16	4.15	9.04	0.59	1.96
42 – Tehama-Glenn-Colusa	3.20	2.66	7.27	1.15	1.76	2.92	2.92	6.07	0.78	1.67
43 – Del Norte-Siskiyou-Lassen-Trinity-Modoc-Plumas-Sierra	3.11	2.89	9.04	1.31	1.73	3.60	3.28	12.14	0.79	1.82
44 – Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	2.43	2.27	5.16	0.71	1.54	2.91	2.62	6.70	1.13	1.67

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Note: Max = maximum DEFF value; Min = minimum DEFF value.

Table 9-2. Design effect (DEFF) and square root DEFF (DEFT) statistics for estimates from the child interviews, overall and by reported stratum within study year

Reported stratum	CHIS 2015					CHIS 2016				
	DEFF				Average DEFT	DEFF				Average DEFT
	Average	Median	Max	Min		Average	Median	Max	Min	
State	2.80	2.78	4.70	2.12	1.67	3.70	3.91	5.22	2.44	1.91
1 – Los Angeles	2.44	2.29	6.34	1.36	1.54	3.03	3.02	5.47	1.71	1.73
2 – San Diego	2.08	2.08	3.75	0.76	1.41	2.19	2.36	3.05	0.83	1.46
3 – Orange	1.89	1.89	2.58	0.62	1.36	3.56	2.64	7.90	0.32	1.79
4 – Santa Clara	2.44	2.29	6.90	0.07	1.49	3.02	3.56	5.05	0.18	1.66
5 – San Bernardino	1.63	1.59	3.19	0.45	1.26	2.14	2.31	4.19	0.62	1.41
6 – Riverside	2.39	2.18	5.51	0.61	1.49	3.23	2.92	8.04	0.56	1.70
7 – Alameda	1.98	1.99	3.57	0.13	1.37	2.06	2.29	3.91	0.20	1.39
8 – Sacramento	1.88	1.84	5.47	0.23	1.31	2.33	1.95	5.56	0.55	1.46
9 – Contra Costa	1.43	1.37	3.28	0.44	1.17	2.79	2.88	5.95	0.02	1.56
10 – Fresno	1.84	1.94	4.54	0.75	1.32	2.20	2.60	4.28	0.18	1.39
11 – San Francisco	2.40	2.85	4.27	0.66	1.49	1.50	1.09	3.33	0.23	1.16
12 – Ventura	1.67	1.75	2.66	0.16	1.25	2.12	2.34	4.79	0.19	1.33
13 – San Mateo	1.53	1.28	2.90	0.02	1.18	1.89	2.14	3.88	0.06	1.24
14 – Kern	1.76	1.43	3.62	0.81	1.30	2.29	1.89	4.93	0.52	1.44
15 – San Joaquin	1.25	1.24	2.15	0.60	1.11	1.79	1.83	2.71	0.67	1.32
16 – Sonoma	2.21	2.36	3.02	0.42	1.46	1.55	0.95	3.47	0.50	1.17
17 – Stanislaus	1.32	1.40	1.83	0.37	1.13	1.56	1.38	3.17	0.45	1.23
18 – Santa Barbara	1.13	1.22	2.04	0.34	1.03	1.47	1.63	2.28	0.36	1.18
19 – Solano	1.45	1.69	2.18	0.26	1.17	1.25	1.17	2.59	0.18	1.05
20 – Tulare	1.23	1.30	2.01	0.22	1.09	1.48	1.63	2.43	0.36	1.19
21 – Santa Cruz	1.49	1.52	2.98	0.47	1.18	2.19	2.18	4.57	0.39	1.42
22 – Marin	1.59	1.46	2.80	0.64	1.24	1.16	1.11	2.21	0.59	1.05
23 – San Luis Obispo	1.30	1.29	3.00	0.08	1.09	1.78	1.67	4.99	0.26	1.28
24 – Placer	1.39	1.46	2.48	0.09	1.15	1.54	1.55	3.66	0.07	1.18

(continued)

Table 9-2. Design effect (DEFF) and square root DEFF (DEFT) statistics for estimates from the child interviews, overall and by reported stratum within study year (continued)

Reported stratum	CHIS 2015					CHIS 2016				
	DEFF				Average DEFT	DEFF				Average DEFT
	Average	Median	Max	Min		Average	Median	Max	Min	
25 – Merced	1.74	1.48	4.81	0.60	1.28	2.75	3.20	4.61	0.51	1.60
26 – Butte	1.37	1.03	2.72	0.34	1.13	1.78	1.32	4.28	0.27	1.25
27 – Shasta	1.63	1.59	2.12	1.18	1.27	1.77	2.06	2.95	0.22	1.26
28 – Yolo	1.82	1.29	5.90	0.11	1.22	1.85	1.44	5.18	0.56	1.30
29 – El Dorado	1.38	1.20	3.10	0.18	1.07	1.84	1.67	3.93	0.67	1.31
30 – Imperial	1.46	1.47	2.49	0.35	1.18	2.27	2.57	3.51	0.50	1.48
31 – Napa	1.48	1.34	2.12	1.12	1.21	1.47	1.61	3.26	0.19	1.13
32 – Kings	1.52	1.52	2.58	0.19	1.20	1.59	1.66	2.52	0.46	1.23
33 – Madera	1.08	1.10	1.32	0.66	1.04	3.09	3.93	5.54	0.24	1.63
34 – Monterey	2.58	2.13	6.34	0.64	1.54	2.06	1.22	6.29	0.14	1.27
35 – Humboldt	1.16	1.19	1.60	0.69	1.07	1.14	1.03	2.22	0.20	1.03
36 – Nevada	1.39	1.57	1.85	0.40	1.16	1.35	1.29	3.76	0.10	1.09
37 – Mendocino	1.62	1.49	2.74	0.90	1.26	2.19	2.26	3.83	0.56	1.45
38 – Sutter	1.81	1.67	4.78	0.65	1.29	1.30	0.95	3.35	0.44	1.09
39 – Yuba	1.41	1.41	2.49	0.37	1.17	2.32	2.29	4.63	0.18	1.44
40 – Lake	1.35	1.34	2.05	0.47	1.15	1.53	1.28	4.71	0.16	1.14
41 – San Benito	2.38	1.71	5.67	0.07	1.42	1.35	0.97	3.41	0.21	1.11
42 – Tehama-Glenn-Colusa	1.19	0.92	2.16	0.18	1.05	1.33	1.63	2.81	0.10	1.06
43 – Del Norte-Siskiyou-Lassen-Trinity-Modoc-Plumas-Sierra	1.35	1.37	1.77	0.35	1.14	1.26	1.38	3.09	0.06	1.00
44 – Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	1.11	1.18	1.38	0.38	1.04	1.55	1.42	2.76	0.29	1.19

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Note: Max = maximum DEFF value; Min = minimum DEFF value.

Table 9-3. Design effect (DEFF) and square root DEFF (DEFT) statistics for estimates from the teen interviews, overall and by reported stratum within study year

Reported stratum	CHIS 2015					CHIS 2016				
	DEFF				Average DEFT	DEFF				Average DEFT
	Average	Median	Max	Min		Average	Median	Max	Min	
State	3.44	3.24	7.98	0.25	1.81	7.33	6.97	20.07	4.07	2.66
1 – Los Angeles	2.44	2.29	6.34	1.36	1.54	4.85	4.43	9.49	0.99	2.15
2 – San Diego	2.08	2.08	3.75	0.76	1.41	4.85	5.34	7.26	1.64	2.16
3 – Orange	1.89	1.89	2.58	0.62	1.36	2.65	2.72	5.00	0.69	1.57
4 – Santa Clara	2.44	2.29	6.90	0.07	1.49	2.51	3.44	4.37	0.44	1.49
5 – San Bernardino	1.63	1.59	3.19	0.45	1.26	4.02	2.95	10.17	0.14	1.77
6 – Riverside	2.39	2.18	5.51	0.61	1.49	4.77	5.57	7.57	0.54	2.11
7 – Alameda	1.98	1.99	3.57	0.13	1.37	2.25	2.36	4.52	0.00	1.35
8 – Sacramento	1.88	1.84	5.47	0.23	1.31	2.31	1.97	4.48	0.03	1.45
9 – Contra Costa	1.43	1.37	3.28	0.44	1.17	3.77	4.49	6.67	0.01	1.79
10 – Fresno	1.84	1.94	4.54	0.75	1.32	1.44	1.71	2.37	0.12	1.15
11 – San Francisco	2.40	2.85	4.27	0.66	1.49	1.56	1.30	2.82	0.09	1.20
12 – Ventura	1.67	1.75	2.66	0.16	1.25	1.98	2.13	4.94	0.22	1.33
13 – San Mateo	1.53	1.28	2.90	0.02	1.18	1.63	1.38	3.13	0.20	1.20
14 – Kern	1.76	1.43	3.62	0.81	1.30	2.47	2.29	4.58	0.65	1.52
15 – San Joaquin	1.25	1.24	2.15	0.60	1.11	1.63	1.69	3.23	0.17	1.18
16 – Sonoma	2.21	2.36	3.02	0.42	1.46	**	**	**	**	**
17 – Stanislaus	1.32	1.40	1.83	0.37	1.13	1.84	1.98	3.69	0.07	1.19
18 – Santa Barbara	1.13	1.22	2.04	0.34	1.03	2.43	2.85	4.09	0.25	1.48
19 – Solano	1.45	1.69	2.18	0.26	1.17	**	**	**	**	**
20 – Tulare	1.23	1.30	2.01	0.22	1.09	1.51	1.85	3.54	0.07	1.11
21 – Santa Cruz	1.49	1.52	2.98	0.47	1.18	1.29	1.55	2.14	0.26	1.11
22 – Marin	1.59	1.46	2.80	0.64	1.24	1.89	1.62	3.57	0.11	1.31
23 – San Luis Obispo	1.30	1.29	3.00	0.08	1.09	2.57	3.79	3.89	0.21	1.48
24 – Placer	1.39	1.46	2.48	0.09	1.15	1.37	1.48	2.42	0.08	1.11

(continued)

Table 9-3. Design effect (DEFF) and square root DEFF (DEFT) statistics for estimates from the teen interviews, overall and by reported stratum within study year (continued)

Reported stratum	CHIS 2015					CHIS 2016				
	DEFF				Average DEFT	DEFF				Average DEFT
	Average	Median	Max	Min		Average	Median	Max	Min	
25 – Merced	1.74	1.48	4.81	0.60	1.28	1.59	1.18	3.07	0.47	1.20
26 – Butte	1.37	1.03	2.72	0.34	1.13	1.85	2.63	3.04	0.01	1.21
27 – Shasta	1.63	1.59	2.12	1.18	1.27	**	**	**	**	**
28 – Yolo	1.82	1.29	5.90	0.11	1.22	1.42	1.16	3.16	0.78	1.16
29 – El Dorado	1.38	1.20	3.10	0.18	1.07	2.33	2.46	3.71	0.95	1.50
30 – Imperial	1.46	1.47	2.49	0.35	1.18	2.64	2.96	5.38	0.28	1.53
31 – Napa	1.48	1.34	2.12	1.12	1.21	1.82	0.86	4.04	0.02	1.19
32 – Kings	1.52	1.52	2.58	0.19	1.20	1.65	1.96	2.63	0.01	1.21
33 – Madera	1.08	1.10	1.32	0.66	1.04	**	**	**	**	**
34 – Monterey	2.58	2.13	6.34	0.64	1.54	1.62	1.71	3.07	0.01	1.07
35 – Humboldt	1.16	1.19	1.60	0.69	1.07	**	**	**	**	**
36 – Nevada	1.39	1.57	1.85	0.40	1.16	1.24	1.41	1.56	0.67	1.10
37 – Mendocino	1.62	1.49	2.74	0.90	1.26	**	**	**	**	**
38 – Sutter	1.81	1.67	4.78	0.65	1.29	1.90	1.32	3.82	0.01	1.25
39 – Yuba	1.41	1.41	2.49	0.37	1.17	1.04	1.00	2.39	0.05	0.92
40 – Lake	1.35	1.34	2.05	0.47	1.15	1.49	1.39	3.17	0.05	1.08
41 – San Benito	2.38	1.71	5.67	0.07	1.42	1.38	1.67	2.33	0.29	1.13
42 – Tehama-Glenn-Colusa	1.19	0.92	2.16	0.18	1.05	1.41	0.72	3.96	0.11	1.02
43 – Del Norte-Siskiyou-Lassen-Trinity-Modoc-Plumas-Sierra	1.35	1.37	1.77	0.35	1.14	**	**	**	**	**
44 – Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	1.11	1.18	1.38	0.38	1.04	1.47	1.89	2.21	0.01	1.13

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Note: Max = maximum DEFF value; Min = minimum DEFF value.

** = estimates suppressed due to instability linked to small sample size within reported stratum.

9.2 Methods for Variance Estimation

Variance estimation for CHIS comes in two forms. The first is referred to as Taylor Series linearization or *linearization* for short. The analysis weights described in Sections 4-6 along with the design stratum indicator and survey analysis software (e.g., SUDAAN, Stata, SAS/Survey, R) are used to generate (weighted) linearization variance estimates. Design effects (variance given the design divided by the variance under a simple random sample) and coefficients of variation (standard error divided by the estimated average) can be calculated to assess the relative precision of any particular estimate.

The second form of variance estimation is replication. There are several benefits noted for replication variance estimation, including the ability to capture the random nature of the adjustments applied throughout the weighting process. Replicate point estimates (e.g., mean) are generated from replicate weights and used in the following general formula to calculate the associated variance for the point estimate:

$$v\left(\hat{\theta}\right) = a\sum_{r=1}^{R}\left(\hat{\theta}_{(r)} - \hat{\theta}\right)^2 \tag{9.1}$$

where $\hat{\theta}_{(r)}$ is the estimate generated from the *r*th replicate; $\hat{\theta}$ is the full-sample estimate of a specific form that depends on the variance estimator chosen (e.g., estimate generated using the linearization weight); and *a* is a constant depending on the replication method chosen. Replicate weights were formed by first adjusting the base weights for the subsampling and then administering all adjustments applied to the linearization weight to the replicates weights. See Wolter (2007) for a detailed discussion of variance estimation.

CHIS 2015-2016 employed the same methodology as in past rounds of CHIS—a paired-unit grouped jackknife (GJK) replication with R=80 replicates (see, e.g., Valliant et al., 2008). Details of the CHIS replicates are provided in the next section.

9.3 Design of Replicates

Replicate variance estimation requires a set of weights that capture components associated with the sample design and weight adjustments applied to the full-sample weight (Chapters 3-6). The sections below the methods for calculating the replicate weights for the one-year estimates (Section 9.3.1) and the two-year estimates (Section 9.3.2).

9.3.1 One-Year Replicates

A paired jackknife replication method (JK2) was used for computing variances in CHIS 2015-2016 to maintain consistency with prior years of the study. The benefits a replication method include, for example, the ability to reflect all components of the design and the survey weights into the estimates of precision without the need to know such information. For example, Chapters 3-6 detailed a several adjustments applied to the weights to address sampling and subsampling for nonresponse and to limit biases associated with nonresponse and coverage. The replicate weights were constructed to capture variability in the adjustments.

Construction of the JK2 replicate weights follows procedures developed previously for CHIS. A total of 80 replicates were again created to maintain the same degrees of freedom as in CHIS 2013-2014. The basis for the replicates was constructed as follows within each design stratum:

1) Sampled telephone numbers and addresses were listed in the same order as when they were selected by associated sampling frame. Sampled telephone numbers and addresses are referred to as sample units in the discussion below.
2) The ordered sample units were paired within the list and assigned to the variance strata in a circular fashion. Once the 40th pair was assigned to replicate variance stratum 40, the next pair was assigned to variance stratum 1 and so on.
3) Each sample unit pair was randomly assigned to group (replicate variance unit) 1 or group 2 within variance stratum.

The desired result is to have variance strata for each variance unit designation with roughly the same number of sample units and for each variance strata to be a miniature representation of the full sample. In other words, all variance strata should contain sample units from all design strata, all sampling frames, all telephone types (landline, cell, and surname), all sizeable subgroups, and so on.

The replicate weights were then created within each of the 40 strata that contained a random subsample of respondents, nonrespondents, ineligibles and those with unknown eligibility status. The first step was to form the replicate base weights by modifying the household weights shown in (3.3):

$$HW2_i^{(r)} = \begin{cases} 2 \times HW2_i, & \text{if sample unit } i \text{ is in variance stratum } s \\ & \text{and variance (group) unit } v \\ 0, & \text{if sample unit } i \text{ is in variance stratum } s \\ & \text{and not in variance unit } v \\ HW2_i, & \text{if sample unit } i \text{ is in not in variance stratum } s \end{cases} \quad (9.2)$$

where s=1-40 to index the replicate variance strata; v=1-2 to index the replicate variance units; and r indexes the replicate weights calculated as $r=2 \times (s-1) + v$. For example, units in group 2 (v=2) within variance stratum 40 (s=40) have their input weight multiple by two within replicate 80.

The same steps implemented for the full sample (linear) weight discussed in Chapters 3-6 were then applied independently to each replicate base weight. Adjustments for nonresponse were applied using the same model developed for the full sample; modifications to these models were sometimes required for a few replicates because of small sample sizes for certain subgroups (e.g., teen respondents within design stratum). The final step was to calibrate the weights to the DoF population estimates used for the full sample. Thus, the weight sums for the replicates and full sample estimate the size of the CHIS target population and should match apart from rounding or deviations from the full-sample calibration model.

9.3.2 Two-Year Replicates

The creation of the two-year replicate weights followed the same process described in Section 9.3.1. The first replicate from 2015 was combined with the first replicate from 2016 using a composite factor specific to that replicate (i.e., λ_{gj} was re-calculated for each replicate). This was done again for the second replicate and repeated for all replicates from 2015 and 2016, resulting in 80 replicates with two years of respondents each.

As with the Taylor weight, the replicates with two years of respondents were calibrated to the 2016 population totals specific to that interview type using as many uncollapsed levels of the model covariates as possible.

9.4 Software for Computing Variances

As mentioned in Chapter 2 of this report, researchers must account for the CHIS sampling design and use analysis weights to produce design unbiased population estimates. The focus of this section is a

discussion of example software packages to properly accomplish this goal. Choice of software is generally user preference because they produce similar or even equivalent estimates.

- **WesVar, Version 5.1** (Westat, 2007) is provided free of charge from Westat. WesVar is an interactive software program with a graphical interface that includes replication methods to compute variance estimates. Analytic capabilities include descriptive statistics, as well as multivariate linear and logistic regression.

  WesVar requires (1) the identification of the CHIS full (linear) and replicate weights provided on the data file, and (2) the specification of the replication method JK2. This allows the software to properly account for the sample design and the analysis weights.
- **SUDAAN®, Version 11** (RTI, 2012) is software developed by RTI International to analyze correlated data such as those from a survey. Estimated standard errors are available for Taylor series approximation (linearization) or for replication methods. Replication methods are recommended for CHIS to properly account for the complex nature of the analysis weights.

  SUDAAN contains several procedures for analyzing correlated data. For example, descriptive statistics for categorical and continuous variable are calculated with the CROSSTAB and DESCRIPT procedures, respectively. As with WesVar, SUDAAN requires (1) the identification of the CHIS linear weights (WEIGHT statement) and replicate weights (JACKWGTS statement) provided on the data file, and (2) the specification of the replication method using the DESIGN=JACKKNIFE option.
- **SAS®, Version 9.4** (SAS, 2015) also includes various procedures to analyze complex survey data and provide either linearization or replication variance estimates. The latter methodology is invoked with a REPWEIGHTS statement. For example, PROC SURVEYFREQ is used for categorical variables. VARMETHOD=JACKKNIFE requests the appropriate variance estimation method for CHIS.
- **Stata, Version 15** (StataCorp, 2017) is another option for analyzing CHIS data. Stata contains a list of survey procedures accessed via svy commands to analyze data from sample surveys. For example, “svy mean” and “svy total” produce estimated means and totals, respectively. Replication variance estimates are requested with “svyset” by identifying the linear weights with the “pw” option, the replicate weights with the “jkrweight” option, and the design as “vce(jack).”
- **R, Version 3.4.1** (Venables et al., 2017) is the last software commented on in this short discussion. R is a free software and contains several packages that house procedures for

analyzing survey data such as “survey” (Lumley, 2017) and “PracTools” (Valliant et al., 2017). As with the other packages, R will generate either linearization or replication variance estimates for a variety of statistics. Design objects are first specified via the “svydesign” command to define the type of variance estimation required; “svrepdesign” is needed specifically for replication variances. Functions such as “svymean” and “svytable” then operate on the design objects to produce the associated estimates.

Replication variance estimates are recommended. However, the CHIS data files contain two variables that enable calculation of Taylor-series linearization standard errors.

- TSVARSTR (Taylor’s series variance stratum) – identifies the variance strata. This variable was created by sequentially numbering the design strata separately by sampling frame and year. TSVARSTR must be specified in the software packages when linearization standard errors are desired.
- TSVRUNIT (Taylor’s series unit) – identifies the household cluster for those with multiple person interviews. This variable was created by sequentially numbering participating households within design stratum. In contrast to TSVARSTR, TSVRUNIT is needed only for analyses involving multiple respondents per household (adult and child/teen, child and teen, or adult, child and teen).

10. REFERENCES

Andridge, R.R. and Little, R.J.A. (2010). A Review of Hot Deck Imputation for Survey Non-response. *International Statistical Review*, *78*(1): 40–64. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3130338/.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth.

Brick, J.M. and Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, ***5***(3): 215-238.

Blumberg, S.J. and Luke, J.V. (2016). Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July–December 2014. Division of Health Interview Statistics, National Center for Health Statistics. URL: http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201608.pdf.

Chen, P., et al. (2014). National Survey on Drug Use and Health: Person-level sampling, Weight Calibration. Prepared for the Substance Abuse and Mental Health Services Administration as part of the 2012 Methodological Resource Book, February 2014. http://www.samhsa.gov/data/sites/default/files/NSDUH2012MRB-Ammended/NSDUHmrbSamplingWgt2012.pdf

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

Kim, J., Li, J., and Valliant, R. (2007). Cell collapsing in poststratification. *Survey Methodology*, ***33***:139-150.

Kott, P.S. (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. *Survey Methodology*, *32*(2): 133-142.

Levine, B. and Harter, R. (2015). Optimal Allocation of Cell-Phone and Landline Respondents in Dual-Frame Surveys. *e*, *79*(1): 91–104, https://doi.org/10.1093/poq/nfu044.

Lumley, T. (2017). *R Package 'survey': Analysis of Complex Survey Samples* (June 23, 2017), https://cran.r-project.org/web/packages/survey/survey.pdf. Accessed 10 SEP 2017.

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

RTI International. (RTI, 2012). *SUDAAN Language Manual*, Volumes 1 and 2, Release 11. Research Triangle Park, NC: RTI International is a tradename of Research Triangle Institute.

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

StataCorp. (2017). *Stata Statistical Software*: Release 15. College Station, TX: StataCorp LP.

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

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

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

Valliant, R., Brick, J.M., and Dever, J. (2008). Weight adjustments for the grouped jackknife variance estimator. *Journal of Official Statistics*, *24*, 469-488.

Valliant, R., Dever, J.A., and Kreuter, F. (2017). *R Package 'PracTools': Tools for Designing and Weighting Survey Samples* (26 July 2017), https://cran.r-project.org/web/packages/PracTools/PracTools.pdf. Accessed 10 SEP 2017.

Venables, W.N., Smith, D.M., and the R Core Team (2017). *An Introduction to R - Notes on R: A Programming Environment for Data Analysis and Graphics*, Version 3.4.1 (2017-06-30), https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf. Accessed 10 SEP 2017.

Westat. (2007). WesVar ™ 4.0 user's guide. Rockville, MD: Westat.

Wolter, K. M. (2007). *Introduction to Variance Estimation*, 2nd edition. Springer.

Appendix A – Frame Sizes, Sample Sizes, and Base Weights

Appendix A includes supplemental information on the CHIS 2015-2016 sample design directly related to calculation of the base weights (inverse probability of selection).

Table A-1 contains counts for landline and surname sampling frames from the last quarter of CHIS 2015 data collection by design stratum, as well as the associated sample sizes and base weights averaged across quarters. Table A-2 provides the same information for CHIS 2016 as shown in Table A-1 in addition to information for the Imperial Count addressed-based sampling information.

Table A-3 contains cell phone sampling frame counts from the last quarter of CHIS 2015 by design stratum, and the associated sample sizes and base weights averaged across quarters. Table A-4 includes cell phone information for CHIS 2016.

Table A-1. CHIS 2015 Frame count, sample size, and base weight for landline and surname samples by design stratum

Design stratum	Landline Frame			Japanese Surname Frame			Korean Surname Frame			Vietnamese Surname Frame		
	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]
1 – Los Angeles	6,196,702	109,401	444	1,305,967	186	140	1,486,655	2,229	101	1,123,379	592	83
2 – San Diego	3,859,497	37,280	472	61,311	32	109	125,064	291	106	161,753	211	137
3 – Orange	2,562,844	37,559	365	107,833	54	161	185,730	770	98	271,437	658	101
4 – Santa Clara	1,800,575	23,542	379	35,397	43	117	66,070	644	101	137,028	488	124
5 – San Bernardino	1,780,799	17,798	350	72,306	15	124	58,409	193	62	71,706	76	77
6 – Riverside	1,693,863	23,860	336	43,834	15	105	43,591	152	50	49,552	73	58
7 – Alameda	1,197,188	18,904	285	16,825	26	88	28,326	544	41	30,924	210	45
8 – Sacramento	1,088,188	16,791	307	6,584	23	72	23,308	210	34	28,339	131	36
9 – Contra Costa	1,065,407	9,940	378	15,437	18	64	15,197	197	22	12,445	63	24
10 – Fresno	1,687,300	12,365	388	13,080	13	71	13,341	60	27	15,889	23	31
11 – San Francisco	7,841,039	10,831	1,279	3,245,457	18	514	2,592,401	646	377	2,437,461	174	358
12 – Ventura	4,603,915	8,057	905	4,379,678	10	670	2,615,861	67	382	2,163,064	27	320
13 – San Mateo	2,613,965	7,962	959	288,944	17	169	564,003	245	210	461,301	53	177
14 – Kern	2,199,429	5,835	1,054				571,829	25	297	865,568	11	441
15 – San Joaquin	1,962,477	5,432	1,068	1,195	8	75	794,628	54	405	1,035,811	46	524
16 – Sonoma	1,687,938	4,978	1,124				585,621	37	495	1,014,506	16	849
17 – Stanislaus	1,464,825	4,980	888				648,990	25	359	491,515	11	277
18 – Santa Barbara	1,505,179	4,234	869				491,759	22	275	1,013,246	8	552
19 – Solano	1,533,790	5,550	1,097				222,698	33	215	871,393	14	798
20 – Tulare	1,161,217	5,300	843				178,455	11	157			
21 – Santa Cruz	981,230	4,405	836									

(continued)

Table A-1. CHIS 2015 Frame count, sample size, and base weight for landline and surname samples by design stratum (continued)

Design stratum	Landline Frame			Japanese Surname Frame			Korean Surname Frame			Vietnamese Surname Frame		
	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]
22 – Marin	2,015,109	40,857	408									
23 – San Luis Obispo	2,739,169	3,211	534				1,606,030	11	57			
24 – Placer	3,000,015	3,642	513				1,071,320	33	42	2,038,345	13	67
25 – Merced	3,446,354	5,370	289				3,312,033	8	104			
26 – Butte	2,219,563	2,706	654				4,076,314	13	120	250	7	18
27 – Shasta	2,412,723	2,840	611				1,962,770	9	65			
28 – Yolo	3,109,010	3,510	448				1,394,965	19	51	4,415,856	6	129
29 – El Dorado	2,944,914	3,559	449									
30 – Imperial	2,754,884	5,349	285									
31 – Napa	2,405,788	5,173	287				1,104,106	8	41			
32 – Kings	2,986,486	6,691	267									
33 – Madera	3,512,354	4,492	477									
34 – Monterey	3,409,939	7,256	290	821	6	68	631,719	28	33	803,197	11	37
35 – Humboldt	1,241,646	2,343	779									
36 – Nevada	1,295,600	2,949	560									
37 – Mendocino	1,222,188	2,711	619									
38 – Sutter	1,217,032	4,076	326									
39 – Yuba	1,340,428	4,701	293									
40 – Lake	1,042,477	2,894	517									
41 – San Benito	580,907	7,032	132									

(continued)

Table A-1. CHIS 2015 Frame count, sample size, and base weight for landline and surname samples by design stratum (continued)

Design stratum	Landline Frame			Japanese Surname Frame			Korean Surname Frame			Vietnamese Surname Frame		
	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]
42 – Tehama-Glenn-Colusa	971,691	2,922	407									
43 – Del Norte-Siskiyou-Lassen-Trinity- Modoc-Plumas-Sierra	1,107,563	2,675	508				214	8	13			
44 – Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	1,140,319	2,717	501				687,125	12	70	458,062	6	52

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

n = number of telephone numbers

[1] Frame counts for working 100-number blocks provided in the last quarter of data collection for CHIS 2015

[2] Total sample size for full year of data collection

[3] Average base weight (inverse probability of selection). Values in CHIS 2015 compared to CHIS 2016 were higher because of small supplemental samples selected throughout the data collection period.

Table A-2. CHIS 2016 Frame count, sample size, and base weight for landline, surname, and address-based samples by design

Design stratum	Landline Frame			Japanese Surname Frame			Korean Surname Frame			Vietnamese Surname Frame		
	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]
1 – Los Angeles	6,579,806	81,054	85	319,628	4,384	13	497,327	13,879	31	192,584	1,017	23
2 – San Diego	2,538,768	62,348	62	250,836	618	33	285,816	1,527	42	169,660	307	21
3 – Orange	4,300,695	20,064	133	171,099	1,519	21	269,688	5,998	28	131,345	1,423	27
4 – Santa Clara	3,302,130	13,475	118	38,196	1,116	10	146,243	4,441	11	115,612	970	23
5 – San Bernardino	3,183,432	10,403	151	201,633	389	13	166,099	1,312	11	31,593	126	10
6 – Riverside	2,686,139	19,156	115	160,346	378	22	175,158	945	24	67,633	118	15
7 – Alameda	3,065,286	7,660	164	35,145	535	11	223,766	2,926	15	116,621	313	30
8 – Sacramento	771,926	8,428	89	174,531	418	15	265,322	1,013	18	96,176	162	33
9 – Contra Costa	1,081,647	9,579	72	214,731	289	11	194,165	862	10	77,055	71	12
10 – Fresno	1,071,340	7,381	66	137,103	285	10	320,274	332	13	104,187	37	19
11 – San Francisco	1,567,232	6,239	89	186,391	422	13	224,219	3,873	16	78,696	296	23
12 – Ventura	1,043,942	10,594	67	120,307	220	15	170,885	347	15	826	37	7
13 – San Mateo	1,114,004	8,889	75	143,491	427	47	162,927	1,642	55	47,395	87	9
14 – Kern	684,979	9,431	47				230,439	146	23	76,990	26	19
15 – San Joaquin	993,777	4,215	73	193,803	171	20	338,279	276	30	90,135	64	23
16 – Sonoma	867,279	3,158	84				262,550	126	27	99,388	18	27
17 – Stanislaus	878,405	5,689	69				226,645	163	22	43,307	25	12
18 – Santa Barbara	917,002	3,720	94				199,245	118	20	36,051	20	10
19 – Solano	723,508	5,684	53				311,207	153	73	89,831	20	19
20 – Tulare	860,305	4,840	82				250,222	89	57			
21 – Santa Cruz	629,198	6,202	80									

(continued)

Table A-2. CHIS 2016 Frame count, sample size, and base weight for landline, surname, and address-based samples by design (continued)

Design stratum	Landline Frame			Japanese Surname Frame			Korean Surname Frame			Vietnamese Surname Frame		
	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]
22 – Marin	815,512	5,111	71									
23 – San Luis Obispo	696,882	3,788	61				135,686	65	16			
24 – Placer	872,374	4,033	75				187,450	134	24	109,549	19	42
25 – Merced	609,299	4,765	39				243,128	93	35			
26 – Butte	666,868	2,163	68				114,096	59	17	238	13	5
27 – Shasta	646,437	3,017	128				36,859	38	8			
28 – Yolo	698,049	4,260	55				182,146	78	27	99,530	14	21
29 – El Dorado	620,765	4,181	55									
30 – Imperial	665,336	4,447	39									
Imperial (ABS)	5,940	4,180	3									
31 – Napa	781,764	5,660	38				50,080	91	7			
32 – Kings	556,050	6,885	28									
33 – Madera	724,644	3,663	41									
34 – Monterey	975,463	3,753	83	129,748	131	18	148,565	198	20	30,809	27	10
35 – Humboldt	519,863	3,706	44									
36 – Nevada	599,380	3,573	69									
37 – Mendocino	502,894	3,407	41									
38 – Sutter	495,207	5,208	41									
39 – Yuba	651,246	4,018	47									
40 – Lake	574,914	4,171	63									
41 – San Benito	339,135	9,187	25									

(continued)

Table A-2. CHIS 2016 Frame count, sample size, and base weight for landline, surname, and address-based samples by design (continued)

Design stratum	Landline Frame			Japanese Surname Frame			Korean Surname Frame			Vietnamese Surname Frame		
	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]
42 – Tehama-Glenn-Colusa	635,835	2,564	45									
43 – Del Norte-Siskiyou-Lassen-Trinity- Modoc-Plumas-Sierra	1,035,852	2,723	70				15,304	30	4			
44 – Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	615,088	3,721	61				71,584	63	11	28,499	16	8

Source: UCLA Center for Health Policy Research. 2016 California Health Interview Survey.

n = number of telephone numbers; ABS = address-based sampling

[1] Frame counts for working 100-number blocks provided in the last quarter of data collection for CHIS 2015

[2] Total sample size for full year of data collection

[3] Average base weight (inverse probability of selection).

Table A-3. CHIS 2015-2016 Frame count, sample size, and base weight for cell phone samples by design stratum

Design stratum	CHIS 2015 Cell Sample			CHIS 2016 Cell Sample		
	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]
1 – Los Angeles	12,281,149	34,149	1,284	15,505,708	41,852	368
2 – San Diego	4,133,046	14,430	989	4,905,927	33,930	143
3 – Orange	4,005,545	9,918	1,396	4,846,790	11,481	416
4 – Santa Clara	2,360,871	7,851	945	2,790,006	7,421	373
5 – San Bernardino	2,202,193	5,379	1,505	2,789,841	6,075	458
6 – Riverside	2,292,107	9,100	876	2,830,334	11,479	246
7 – Alameda	1,950,241	4,870	1,484	2,517,070	6,196	403
8 – Sacramento	1,381,967	3,773	2,181	2,066,866	4,564	450
9 – Contra Costa	889,580	3,508	1,106	1,210,534	5,885	203
10 – Fresno	1,118,256	3,693	1,115	1,527,375	4,678	325
11 – San Francisco	1,399,722	3,011	1,650	1,770,933	5,429	319
12 – Ventura	960,925	2,413	1,423	1,245,636	3,283	370
13 – San Mateo	689,811	2,992	814	850,677	6,482	130
14 – Kern	988,680	3,193	1,082	1,235,898	3,670	334
15 – San Joaquin	684,328	1,922	1,382	898,021	2,969	300
16 – Sonoma	528,963	1,380	1,731	621,743	1,581	392
17 – Stanislaus	607,954	2,860	639	709,630	3,673	193
18 – Santa Barbara	503,836	2,080	874	569,893	2,820	198
19 – Solano	406,940	1,682	1,063	540,445	2,832	189
20 – Tulare	471,895	1,699	1,099	570,547	4,170	135
21 – Santa Cruz	324,742	2,260	460	320,660	2,697	116
22 – Marin	251,136	30,245	30	390,075	8,098	45
23 – San Luis Obispo	386,482	2,561	462	363,332	4,015	89
24 – Placer	409,126	2,358	576	483,215	3,828	124
25 – Merced	257,746	1,858	538	346,597	3,409	99
26 – Butte	374,808	1,320	836	291,495	1,782	160
27 – Shasta	212,726	2,035	371	250,523	3,694	66
28 – Yolo	351,195	1,657	387	185,781	2,706	67
29 – El Dorado	141,871	2,017	246	154,557	3,574	42

(continued)

Table A-3. CHIS 2015-2016 Frame count, sample size, and base weight for cell phone samples by design stratum (continued)

Design stratum	CHIS 2015 Cell Sample			CHIS 2016 Cell Sample		
	Frame[1] (n)	Sample[2] (n)	Weight[3]	Frame[1] (n)	Sample[2] (n)	Weight[3]
30 – Imperial	304,898	2,658	351	361,004	4,891	73
31 – Napa	123,638	2,707	137	139,490	4,225	31
32 – Kings	151,674	2,594	183	189,801	2,753	65
33 – Madera	150,718	2,294	219	190,166	3,915	47
34 – Monterey	386,421	1,622	1,098	549,581	2,090	263
35 – Humboldt	428,922	1,362	458	183,406	2,273	78
36 – Nevada	97,122	2,193	146	110,298	3,500	30
37 – Mendocino	111,726	2,211	166	127,510	3,573	34
38 – Sutter	242,049	13,655	36	257,519	8,052	30
39 – Yuba [4]	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
40 – Lake	64,266	1,747	121	70,017	2,554	23
41 – San Benito	71,985	3,013	69	92,198	3,430	23
42 – Tehama-Glenn-Colusa	72,819	1,430	185	91,366	2,125	40
43 – Del Norte-Siskiyou-Lassen-Trinity-Modoc-Plumas-Sierra	1,813,943	142,021	168	158,820	2,023	76
44 – Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	144,819	1,386	428	200,194	3,509	57

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

n = number of telephone numbers; *N/A* = not applicable

[1] Frame counts for working 100-number blocks provided in the last quarter of data collection for CHIS 2015

[2] Total sample size for full year of data collection

[3] Average base weight (inverse probability of selection). Values in CHIS 2015 compared to CHIS 2016 were higher because of small supplemental samples selected throughout the data collection period.

[4] Yuba county does not contain a cell phone rate center and therefore was not available for sampling.

Appendix B – Summary Statistics for Weights and Weight Adjustments

Appendix B includes summary statistics on the CHIS 2015-2016 base weights, analysis weights, and the weight adjustments by person interview (adult, child and teen).

Table B-1 contains summary statistics for the household weight (Chapter 3) used as the basis for the person-level weights.

Table B-2, Table B-3, and Table B-4 includes summary information for the adult weights (Chapter 4), child weights (Chapter 5) and teen weights (Chapter 6).

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

Survey Weight Statistics[1]	CHIS 2015 Sampling Frame			CHIS 2016 Sampling Frame			
	Landline	List	Cell	Landline	List	Cell	ABS
1. Base weight							
1.1 Sample size	271,032	8,274	206,429	188,622	70,632	253,186	4,180
1.2 Sum of weights	65,436,913.0	324,273.3	211,735,600.0	59,998,912.6	7,944,524.3	220,520,898.3	13,196.6
1.3 Coefficient of variation	61.7	62.5	75.9	97.4	200.9	91.2	34.8
2. Adjustment for multiple periods and frames							
2.1 Sample size	271,032	8,274	206,429	188,622	70,632	253,186	4,180
2.2 Sum of weights	17,738,575.6	43,998.7	52,923,300.0	15,079,061.6	1,979,015.2	54,867,000.0	13,196.6
2.3 Coefficient of variation	61.2	62.3	74.6	98.2	199.0	91.0	34.8
3. Phase 2 subsampling adjustment							
3.1 Sample size	225,200	5,117	158,262	46,858	38,734	137,231	0
3.2 Sum of weights	18,049,934.9	45,494.5	53,866,475.6	16,038,102.5	2,208,573.5	55,726,221.7	13,196.6
3.3 Coefficient of variation	84.3	98.2	108.9	161.8	290.8	126.2	34.8
3.4 Mean non-zero adjustment factor	1.2	1.7	1.3	1.2	1.6	1.5	1.0
4. Unknown residential status adjustment							
4.1 Sample size							
a. Known residential status	25,063	1,040	34,328	138,527	45,448	135,299	4,180
b. Unknown residential	200,137	4,077	123,934	50,095	25,184	117,887	0
4.2 Sum of weights	2,586,783.5	9,541.9	19,545,237.9	13,620,403.3	1,779,562.2	55,726,221.7	13,196.6
4.3 Coefficient of variation	85.7	97.8	112.9	153.0	284.2	116.4	34.8
4.4 Mean non-zero adjustment factor	1.1	1.1	1.5	1.2	1.2	1.8	1.0

(continued)

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

Survey Weight Statistics[1]	CHIS 2015 Sampling Frame			CHIS 2016 Sampling Frame			
	Landline	List	Cell	Landline	List	Cell	ABS
5. Unknown eligibility adjustment							
5.1 Sample size							
a. Residential, Eligible	24,389	1,026	21,521	14,757	8,821	23,785	4,180
b. Residential, Unknown eligibility	674	14	12,807	20,695	17,809	1,056	0
5.2 Sum of weights	2,586,593.1	9,731.8	19,545,237.9	3,722,191.6	953,829.7	12,392,021.0	13,196.6
5.3 Coefficient of variation	85.1	94.7	91.9	155.8	232.5	105.2	34.8
5.4 Mean non-zero adjustment factor	1.0	1.1	2.0	2.1	2.3	1.0	1.0
6. Screener nonresponse adjustment							
6.1 Sample size							
a. Screener respondents	22,137	937	16,620	13,170	7,482	19,531	500
b. Screener nonrespondents	2,252	89	4,901	1,587	1,339	4,254	3,680
6.2 Sum of weights (screener respondents)	2,586,590.9	9,731.8	19,545,237.9	3,722,191.6	953,829.7	12,392,021.0	13,196.6
6.3 Coefficient of variation	82.3	91.7	88.0	153.6	232.3	102.2	21.0
6.4 Mean non-zero adjustment factor	1.2	1.2	1.3	1.1	1.1	1.2	9.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

[1] Statistics after application of each weight adjustment

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

Survey Weight Statistics[1]	CHIS 2015 Sampling Frame			CHIS 2016 Sampling Frame			
	Landline	List	Cell	Landline	List	Cell	ABS
1. Adult Base weight							
1.1 Sample size	22,138	937	16,620	13,170	7,482	19,531	500
1.2 Sum of weights	4,961,093.0	20,197.9	19,696,483.6	6,975,128.4	2,057,677.1	12,487,295.8	25,611.1
1.3 Coefficient of variation	98.8	112.6	89.6	172.0	247.3	104.2	39.7
2. Phase 2 Adjustment							
2.1 Sample size	22,138	937	16,620	13,170	7,482	19,531	500
2.2 Sum of weights	4,910,492.4	19,886.8	19,264,756.3	6,920,864.2	2,026,491.7	12,266,651.3	25,611.1
2.3 Coefficient of variation	99.0	114.1	89.7	172.4	248.8	104.8	39.7
3. Nonresponse adjustment							
3.1 Sample size							
a. Adult respondents	11,211	463	9,360	7,100	3,413	10,284	258
b. Adult nonrespondents	10,927	474	7,260	6,070	4,069	9,247	242
3.2 Sum of weights	4,907,491.7	22,887.6	19,264,756.3	7,155,873.7	1,791,482.2	12,266,651.3	25,611.1
3.3 Coefficient of variation	120.7	117.3	89.3	205.0	248.5	108.4	60.0
3.4 Mean non-zero adjustment factor	2.2	2.6	2.0	2.1	2.4	2.1	2.0
4. Calibration to telephone service							
4.1 Sample size	11,211	463	9,360	7,100	3,413	10,284	258
4.2 Sum of weights	6,330,625.0	30,351.9	21,837,004.0	12,441,885.0	2,986,269.4	27,587,040.5	25,611.1
4.3 Coefficient of variation	126.8	113.7	90.0	211.2	250.1	113.3	60.0
4.4 Mean non-zero adjustment factor	0.8	0.8	0.7	1.7	1.7	2.3	1.0

(continued)

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

Survey Weight Statistics[1]	CHIS 2015 Sampling Frame			CHIS 2016 Sampling Frame			
	Landline	List	Cell	Landline	List	Cell	ABS
5. Composite (phone frames) weight							
5.1 Sample size	11,211	463	9,360	7,100	3,413	10,284	258
5.2 Sum of weights	5,053,029.1	24,739.6	15,135,887.5	7,723,649.0	1,962,796.7	19,737,268.8	25,611.1
5.3 Coefficient of variation	138.9	118.8	103.1	200.2	247.4	149.8	60.0
5.4 Mean non-zero adjustment factor	1.0	1.0	1.0	0.7	0.7	0.7	1.0
6. Composite (phone and ABS) weight [2]							
6.1 Sample size	*N/A*	*N/A*	*N/A*	7,100	3,413	10,284	258
6.2 Sum of weights (screener respondents)	*N/A*	*N/A*	*N/A*	7,721,796.6	1,960,657.0	19,710,697.8	21,523.3
6.3 Coefficient of variation	*N/A*	*N/A*	*N/A*	200.2	247.7	150.0	60.0
6.4 Mean non-zero adjustment factor	*N/A*	*N/A*	*N/A*	1.0	1.0	1.0	0.8
7. Calibration adjustment							
7.1 Number of Trimmed Records	0	0	4	0	0	1	0
7.2 Sum of weights	9,270,736.6	67,053.2	19,744,862.3	7,070,797.2	2,125,864.8	20,165,606.6	27,931.2
7.3 Coefficient of variation	154.9	133.9	111.2	216.7	290.1	161.7	95.7
7.4 Mean non-zero adjustment factor	1.9	3.0	1.7	0.9	1.1	1.1	1.2
7.5 Mean Weight	826.9	144.8	2,109.5	995.9	622.9	1,960.9	108.3

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

N/A = not applicable.

[1] Statistics after application of each weight adjustment.

[2] CHIS 2015 did not include an address-based sample.

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

Survey Weight Statistics[1]	CHIS 2015 Sampling Frame			CHIS 2016 Sampling Frame			
	Landline	List	Cell	Landline	List	Cell	ABS
1. Adult nonresponse adjustment							
1.1 Sample size							
a. Adult respondents	1,011	47	1,720	427	360	2,129	40
b. Adult nonrespondents	1,557	38	1,455	251	236	225	25
1.2 Sum of weights	868,433.6	2,574.2	5,678,588.1	478,153.7	231,141.9	1,557,069.0	3,771.9
1.3 Coefficient of variation	99.8	80.8	92.6	189.2	242.2	115.5	30.9
1.4 Mean non-zero adjustment factor	1.6	1.5	2.1	1.1	1.1	1.1	0.0
2. Base weight							
2.1 Sample size	2,568	85	3,175	678	596	2,354	65
2.2 Sum of weights	945,910.4	2,233.8	6,337,679.7	434,405.0	185,924.6	1,199,610.5	3,224.2
2.3 Coefficient of variation	105.8	68.7	107.7	381.7	399.8	129.1	75.2
3. Nonresponse adjustment							
3.1 Sample size							
a. Child interview respondents	997	36	1,124	386	307	1,412	31
b. Child interview nonrespondents	1,571	49	2,051	292	289	942	34
3.2 Sum of weights	945,910.4	2,233.8	6,337,679.7	434,405.0	185,924.6	1,199,610.5	3,224.2
3.3 Coefficient of variation	111.8	82.2	104.0	339.1	371.8	128.3	105.1
3.4 Mean non-zero adjustment factor	2.2	2.1	2.3	1.4	1.8	1.7	1.9
4. Calibration to telephone service							
4.1 Sample size	997	36	1,124	386	307	1,412	31
4.2 Sum of weights	942,271.6	2,115.3	4,858,229.6	1,862,801.6	810,721.9	5,832,889.0	3,224.2
4.3 Coefficient of variation	123.5	83.5	104.7	335.9	363.5	139.7	105.1
4.4 Mean non-zero adjustment factor	1.0	1.0	0.8	4.3	4.4	4.9	1.0

(continued)

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

Survey Weight Statistics[1]	CHIS 2015 Sampling Frame			CHIS 2016 Sampling Frame			
	Landline	List	Cell	Landline	List	Cell	ABS
5. Composite (phone frames) weight							
5.1 Sample size	997	36	1,124	386	307	1,412	31
5.2 Sum of weights	736,802.1	1,603.3	3,524,259.8	1,105,478.3	509,795.4	4,407,415.7	3,224.2
5.3 Coefficient of variation	139.4	91.3	119.3	325.3	341.4	175.0	105.1
5.4 Mean non-zero adjustment factor	0.7	0.7	0.8	0.6	0.6	0.7	1.0
6. Composite (phone and ABS) weight [2]							
6.1 Sample size	*N/A*	*N/A*	*N/A*	386	307	1,412	31
6.2 Sum of weights (screener respondents)	*N/A*	*N/A*	*N/A*	1,105,018.2	508,918.4	4,402,709.0	2,630.3
6.3 Coefficient of variation	*N/A*	*N/A*	*N/A*	325.4	342.0	175.2	105.1
6.4 Mean non-zero adjustment factor	*N/A*	*N/A*	*N/A*	1.0	1.0	1.0	0.8
7. Calibration adjustment							
7.1 Number of Trimmed Records	0	0	6	0	0	1	0
7.2 Sum of weights	1,584,052.9	5,764.6	4,465,051.1	1,107,683.4	493,408.2	4,425,224.1	5,504.2
7.3 Coefficient of variation	122.7	114.9	105.4	206.0	214.7	156.9	105.0
7.4 Mean non-zero adjustment factor	2.8	4.1	1.7	1.3	1.5	1.1	2.2
7.5 Mean Weight	1,588.8	160.1	3,972.5	2,869.6	1,607.2	3,134.0	177.6

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

N/A = not applicable.

[1] Statistics after application of each weight adjustment.

[2] CHIS 2015 did not include an address-based sample.

Table B-4. Extended interview weighting procedures for teen interviews by sample type

Survey Weight Statistics[1]	CHIS 2015 Sampling Frame			CHIS 2016 Sampling Frame			
	Landline	List	Cell	Landline	List	Cell	ABS
1. Base weight							
1.1 Sample size	2,104	105	1,964	700	637	1428	51
1.2 Sum of weights	646,279.7	2,664.3	2,664,287.2	545,893.2	267,180.7	995,342.9	4,159.2
1.3 Coefficient of variation	111.3	121.7	103.8	220.8	217.2	119.9	73.2
2. Teen nonresponse adjustment							
2.1 Sample size							
a. Teen respondents	367	20	367	230	169	429	12
b. Teen nonrespondents	1,737	85	1,597	470	468	999	39
2.2 Sum of weights	646,279.7	2,664.3	2,664,287.2	545,893.2	267,180.7	995,342.9	4,159.2
2.3 Coefficient of variation	112.8	64.4	130.4	183.1	242.7	145.9	26.7
2.4 Mean non-zero adjustment factor	6.3	5.6	6.0	4.9	6.4	4.4	8.2
3. Calibration to telephone service							
3.1 Sample size	367	20	367	230	169	429	12
3.2 Sum of weights	447,509.6	1,854.1	2,391,396.0	905,631.2	431,808.0	2,823,602.5	4,159.2
3.3 Coefficient of variation	111.2	68.1	114.7	181.4	232.0	170.2	26.7
3.4 Mean non-zero adjustment factor	0.7	0.7	0.9	1.7	1.7	2.9	1.0

(continued)

Table B-4. Extended interview weighting procedures for teen interviews by sample type (continued)

Survey Weight Statistics[1]	CHIS 2015 Sampling Frame			CHIS 2016 Sampling Frame			
	Landline	List	Cell	Landline	List	Cell	ABS
4. Composite (phone frames) weight							
4.1 Sample size	367	20	367	230	169	429	12
4.2 Sum of weights	334,854.8	1,358.8	1,769,994.8	540,745.7	287,651.4	2,133,304.3	4,159.2
4.3 Coefficient of variation	112.6	64.5	141.3	173.2	261.9	214.4	26.7
4.4 Mean non-zero adjustment factor	0.8	0.8	0.7	0.6	0.6	0.6	1.0
5. Composite (phone and ABS) weight [2]							
5.1 Sample size	*N/A*	*N/A*	*N/A*	230	169	429	12
5.2 Sum of weights	*N/A*	*N/A*	*N/A*	540,644.8	287,529.8	2,128,809.9	2,935.9
5.3 Coefficient of variation	*N/A*	*N/A*	*N/A*	173.2	262.0	214.9	26.7
5.4 Mean non-zero adjustment factor	*N/A*	*N/A*	*N/A*	1.0	1.0	1.0	0.7
6. Calibration adjustment							
6.1 Number of Trimmed Records	2	0	8	0	0	0	0
6.2 Sum of weights	831,206.0	6,663.8	2,116,173.8	544,483.4	315,296.3	2,088,112.9	2,664.3
6.3 Coefficient of variation	140.8	128.8	126.3	223.5	233.3	205.7	64.5
6.4 Mean non-zero adjustment factor	2.7	5.3	2.0	1.0	3.6	1.2	0.9
6.5 Mean Weight	2,264.9	333.2	5,766.1	2,367.3	1,865.7	4,867.4	222.0

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

N/A = not applicable.

[1] Statistics after application of each weight adjustment.

[2] CHIS 2015 did not include an address-based sample.