October 2019

CHIS 2017-2018 Methodology Report Series

Report 5

Weighting and Variance Estimation

CALIFORNIA HEALTH INTERVIEW SURVEY

<u>CHIS 2017-2018 METHODOLOGY SERIES</u>

REPORT 5

WEIGHTING AND VARIANCE ESTIMATION

OCTOBER 2019

This report was prepared for the California Health Interview Survey by David Dutwin, Susan Sherr, Arina Goyle, Kathy Langdale, and Jonathan Best of SSRS.

www.chis.ucla.edu

This report describes the weighting and variance estimation methods used in CHIS 2017-2018. 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 2017-2018 Methodology Series: Report 5 - Weighting and Variance Estimation*. Los Angeles, CA: UCLA Center for Health Policy Research, 2019.

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 2017-2018 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 Wellness Foundation, First 5 California, Kaiser Permanente, San Diego County Health and Human Services Agency, Imperial County Public Health Department, UC San Francisco, the Stupski Foundation, California Rural Indian Health Board, and San Francisco Department of Public Health.

PREFACE

Weighting and Variance Estimation is the fifth and final in a series of methodological reports describing the 2017-2018 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, and the Department of Health Care Services. SSRS was responsible for data collection and the preparation of five methodological reports from the 2017-2018 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.

Methodological Report Series for CHIS 2017-2018

The methodological reports for CHIS 2017-2018 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 2017-2018. 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://www.chis.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 2017-2018 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 2017-2018.

- 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://www.chis.ucla.edu/chis/design/Pages/methodology.aspx

The CHIS is a population-based telephone survey of California's residential, noninstitutionalized 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 California 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 2017-2018 data were collected between June 2017 and January 2019. 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 the two year period for the age group corresponding to the data file in use (adult, adolescent, or child). In CHIS 2017-2018, data users will be able to produce single-year estimates using the weights provided (referred to as CHIS 2017 and CHIS 2018, respectively).

See what's new in the 2017-2018 CHIS sampling and data collection here: http://www.chis.ucla.edu/chis/design/Documents/whats-new-chis-2017-2018.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://www.chis.ucla.edu/chis/analyze/Pages/sample-code.aspx.

Additional documentation on constructing the CHIS sampling weights is available in the *CHIS 2017-2018 Methodology Series: Report 5—Weighting and Variance Estimation* posted at http://www.chis.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 2017-2018 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 127 Korean, and 214 Vietnamese adult interviews based on self-identified ethnicity for the 2017-2018 survey year.[1] Additional samples from both the landline and cell phone frames produced 1,375 interviews in 2017-2018 within San Diego County. In 2018, an oversample of American Indian and Alaska Native residents of California added 317 completed interviews, and specific gender and ethnic oversamples in San Francisco provided an additional 498 interviews. Furthermore, an address-based sample from the USPS Delivery Sequence File produced 339 landline or cell phone interviews in 2017 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 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.

---

[1] For the 2017-2018, RDD landline and cell sample frames produced totals of 290 Korean, and 235 Vietnamese adult interviews.

Table 1-1. California county and county group strata used in the CHIS 2017-2018 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, Glenn, Tehama
3. Orange	23. San Luis Obispo	43. Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity
4. Santa Clara	24. Placer	
5. San Bernardino	25. Merced	44. Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne
6. Riverside	26. Butte	

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

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 years 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.

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 2017-2018, 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 880 teen interviews and 3,186 child interviews were completed in CHIS 2017-2018 with approximately 48% of teen interviews and 65% of child interviews 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 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. In 2017-2018, we additionally sampled out-of-area cell phone numbers. These are cell phone numbers with exchanges outside of California that can be matched to an address that is within California, indicating that the owner of the cell phone resides in California but purchased a cell phone in another state.

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.

[2] https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201906.pdf

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 2017-2018.

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.

SSRS designed the methodology and collected data for CHIS 2017-2018, under contract with the UCLA Center for Health Policy Research. SSRS is an independent research firm that specializes in innovative methodologies, optimized sample designs, and reaching low-incidence populations. For all sampled households, SSRS 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 2017-2018 by the type of sample (landline RDD, surname list, cell RDD, and ABS). Note that these figures were accurate as of data collection completion for 2017-2018 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://www.chis.ucla.edu/chis/design/Pages/sample.aspx.

Table 1-2. Number of completed CHIS 2017-2018 interviews by type of sample and instrument

Type of sample[1]	Adult[2]	Child	Adolescent
Total all samples	42,330	3,186	880
Landline RDD[3]	18,896	1,049	434
Cell RDD	21,554	1,996	409
Vietnamese surname list landline	188	10	5
Vietnamese surname list cell phone	80	10	3
Korean surname list landline	354	16	3
Korean surname list cell phone	56	5	1
Both Korean and Vietnamese landline	48	1	1
Imperial County ABS Oversample	339	42	15
AIAN Oversample landline	130	10	-
AIAN Oversample cell phone	187	20	3
San Francisco Oversample landline	148	4	1
San Francisco Oversample cell phone	350	23	5

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

[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. For example, only 157 of the 190 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] Breakdown of completes by frame deviates slightly from original sample numbers due to numbers changing frames following post-sampling database processing.

Interviews in all languages were administered using SSRS's computer-assisted telephone interviewing (CATI) system. The average adult interview took about 42 minutes to complete. The average child and adolescent interviews took about 19 minutes and 24 minutes, respectively. For "child-first" interviews, additional household information asked as part of the child interview averaged about 14 minutes. Interviews in non-English languages typically took longer to complete with an average length of about 50 minutes for the adult interview, 29 minutes for the teen, and 23 minutes for the child. More than eight percent of the adult interviews were completed in a language other than English, as were about 13 percent of all child (parent proxy) interviews and six percent of all adolescent interviews.

Table 1-3 shows the major topic areas for each of the three survey instruments (adult, child, and adolescent). If questions were asked in only one year of survey implementation, the specific year is indicated in the table.

Table 1-3. CHIS 2017-2018 survey topic areas by instrument

Health status	**Adult**	**Teen**	**Child**
General health status	✓	✓	✓
Days missed from work or school due to health problems		✓	✓
Health conditions	**Adult**	**Teen**	**Child**
Asthma	✓	✓	✓
Diabetes, gestational diabetes, pre-diabetes/borderline diabetes	✓		
Heart disease, high blood pressure	✓		
Physical disability	✓		
Physical, behavioral, and/or mental conditions			✓
Developmental assessment, referral to a specialist by a doctor			✓
Mental health	**Adult**	**Teen**	**Child**
Mental health status	✓	✓	
Perceived need, access and utilization of mental health services	✓	✓	
Functional impairment, stigma, three-item loneliness scale (2017)	✓		
Suicide ideation and attempts	✓	✓	
Health behaviors	**Adult**	**Teen**	**Child**
Dietary and water intake, breastfeeding (younger than 3 years)	✓	✓	✓
Physical activity and exercise		✓	✓
Commute from school to home		✓	✓
Walking for transportation and leisure (2017)	✓		
Alcohol, cigarette, and E-cigarette use	✓	✓	
Marijuana use	✓	✓	
Opioid use	✓		
Chewing tobacco, tobacco flavors (2018)	✓	✓	
Exposure to second-hand smoke (2018)	✓		
Sexual behaviors	✓	✓	
HIV testing, HIV prevention medication	✓	✓	
Sleep and technology		✓	
Sedentary time		✓	✓
Contraceptive use	✓	✓	

(continued)

Table 1-3. CHIS 2017-2018 survey topic areas by instrument (continued)

Women's health	**Adult**	**Teen**	**Child**
Pregnancy status, postpartum care	✓		
Dental health	**Adult**	**Teen**	**Child**
Last dental visit, main reason haven't visited dentist	✓	✓	✓
Current dental insurance coverage	✓		✓
Condition of teeth	✓	✓	
Neighborhood and housing	**Adult**	**Teen**	**Child**
Safety, social cohesion	✓	✓	✓
Homeownership	✓		
Length of time at current residence (2017)	✓		
Park use, park and neighborhood safety		✓	✓
Civic engagement	✓	✓	
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)	✓	✓	✓
Communication problems with doctor	✓		✓
Discrimination (2017)	✓		
Timely appointment	✓	✓	✓
Access to specialist and general doctors	✓		
Tele-medical care	✓		
Care coordination (2018)	✓	✓	✓
Voter engagement	**Adult**	**Teen**	**Child**
Voter engagement	✓		
Food environment	**Adult**	**Teen**	**Child**
Access to fresh and affordable foods	✓		
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 assessment of plan	✓	✓	✓
Whether employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months, reasons for lack of insurance	✓	✓	✓
High deductible health plans	✓	✓	✓
Partial scope Medi-Cal	✓		
Medical debt, hospitalizations	✓		

(continued)

Table 1-3. CHIS 2017-2018 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, child support, Social security/pension	✓		
Medi-Cal eligibility, Medi-Cal renewal	✓		
Reason for Medi-Cal non-participation	✓	✓	✓
Bullying	**Adult**	**Teen**	**Child**
Bullying, school safety		✓	
Parental involvement/adult supervision	**Adult**	**Teen**	**Child**
Parental involvement			✓
Parental support, teach support		✓	
Child care and school	**Adult**	**Teen**	**Child**
Current child care arrangements			✓
Paid child care	✓		
First 5 California: Talk, Read, Sing Program / Kit for New Parents			✓
Preschool/school attendance, school name		✓	✓
Preschool quality			✓
School instability, school programs and organizational involvement		✓	
Employment	**Adult**	**Teen**	**Child**
Employment status, spouse's employment status	✓		
Hours worked at all jobs	✓		
Industry and occupation, firm size	✓		
Income	**Adult**	**Teen**	**Child**
Respondent's and spouse's earnings last month before taxes	✓		
Household income, number of persons supported by household income	✓		
Placement on quality of life ladder (2018)	✓		
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, gender identity	✓		
Gender expression		✓	
Living with parents	✓		
Education, English language proficiency	✓		
Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home	✓	✓	✓

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

1.5 Response Rates

The overall response rates for CHIS 2017-2018 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 2017-2018, the landline/list sample household response rate was 5.6 percent (the product of the screener response rate of 10.8 percent and the extended interview response rate at the household level of 52.0 percent). The cell sample household response rate was 3.5 percent, incorporating a screener response rate of 7.1 percent and household-level extended interview response rate of 49.0 percent. CHIS uses AAPOR response rate RR4 (see more detailed in *CHIS 2017-2018 Methodology Series: Report 4 – Response Rates*).

Within the landline and cell phone sampling frames for 2017-2018, the extended interview response rate for the landline/list sample varied across the adult (43.8 percent), child (60.0 percent) and adolescent (25.6 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 40.9 percent, the child rate was 57.5 percent, and the adolescent rate was 18.0 percent in 2017-2018 (see Table 1-4a). 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). 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 2017-2018 response rates – Conditional

Type of Sample	Screener	Household	Adult (given screened)	Child (given screened & eligibility)	Adolescent (given screened & permission)
Overall	8.0%	49.9%	42.3%	58.3%	21.3%
Landline RDD/List	10.8%	52.0%	43.8%	60.0%	25.6%
Cell RDD/List	7.1%	49.0%	40.9%	57.5%	18.0%

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.
Note. This table does not include the Imperial County, AIAN, and San Francisco oversamples.

Table 1-4b. CHIS 2017-2018 response rates – Unconditional

Type of Sample	Screener	Household	Adult (given screened)	Child (given screened & eligibility	Adolescent (given screened & permission)
Overall	8.0%	4.0%	3.4%	4.6%	1.7%
Landline RDD/List	10.8%	5.6%	4.7%	6.4%	2.8%
Cell RDD/List	7.1%	3.5%	2.9%	4.1%	1.3%

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.
Note. This table does not include the Imperial County, AIAN, and San Francisco oversamples

To maximize the response rate, especially at the screener stage, an advance letter in six 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 39.1 percent of the landline RDD sample telephone numbers not identified by the sample vendor as business numbers or not identified by SSRS's dialer software as nonworking numbers, and for 100 percent of surname list sample numbers. Combining these two frames, advance letters were sent to 41.0 percent of all fielded landline telephone numbers. From the onset of 2017 fielding until April of 2018, cell phone sample with matched telephone numbers also received an advance letter. However, after a randomized experiment confirmed that the accuracy of the matching for cell phone sample did not warrant continuing these mailings, they were discontinued (for full experiment details, see Section 7.1 in *CHIS 2017-2018 Methodology Series: Report 4 – Response Rates*). Overall, across the two years, for cell sample, an advance letter was mailed for 27.2 percent of the RDD sample telephone numbers not identified by the sample vendor as business numbers or not identified by SSRS's dialer software as nonworking numbers, and for 100 percent of surname list sample numbers. Combining these two frames, advance letters were sent to 30.4 percent of all fielded cell telephone numbers. As in all CHIS cycles since CHIS 2005, a $2 bill was included with the CHIS 2017-2018 advance letter to encourage cooperation. Unlike previous cycles, additional incentives were not offered to cell phone and nonresponse follow up (NRFU) respondents.

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 CHIS 2017-2018, either a spouse/partner or adult child completed a proxy interview for 20 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://www.chis.ucla.edu/chis/design/Pages/data-quality.aspx.

1.6 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 noninstitutionalized population for each sampling stratum and statewide. The weighting procedures used for CHIS 2017-2018 accomplish the following objectives:

- Compensate for differential probabilities of selection for phone numbers (households) and persons within household;
- Reduce biases occurring because non respondents may have different characteristics than respondents;
- Adjust, to the extent possible, for undercoverage in the sampling frames and in the conduct of the survey; and
- Reduce the variance of the estimates by using auxiliary information

As part of the weighting process, a household weight was created for all households that completed the screener interview. This household weight is the product of the “base weight” (the inverse of the probability of selection of the telephone number) and a variety of adjustment factors. The household weight 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 2017-2018 were created primarily from the California Department of Finance’s (DOF) 2017 and 2018 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 2017-2018 weights are based on 2010 Census counts, as were those used for the 2015-2016 cycle. Please pay close attention when comparing estimates using CHIS 2017-2018 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.

1.7 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. SSRS imputed missing values for those variables used in the weighting process and UCLA-CHPR staff imputed values for nearly every other variable.

Three different imputation procedures were used by SSRS 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 non respondents 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. SSRS 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). The last technique was external data assignment. This method was used for geocoding variables such as strata, Los Angeles SPA, San Diego HSR, and zip where the respondent provided inconsistent information. For such cases geocoding information was used for imputation.

UCLA-CHPR imputed missing values for nearly every variable in the data files other than those imputed by SSRS and some sensitive variables for which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2017-2018 were low, with most variables missing valid responses for less than 1% of the sample. Questions that go to fewer overall respondents or that ask about more sensitive topics can have higher nonresponse.

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, educational attainment and region of California were imputed by SSRS. UCLA-CHPR began their imputation process by imputing household income so that this characteristic was 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 2018 one-year and 2017-2018 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 2017-2018 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. This chapter concludes in Section 2.4 with a brief discussion of quality assurance procedures.

2.1 Weighting Approach

The weighting approach used for CHIS 2017-2018 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 and 2015-2016, an address-based sample (ABS) was selected to supplement the sample of telephone numbers. In CHIS 2017, an ABS sample was chosen to target northern 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 all CHIS person-level interviews: 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, 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 2017-2018 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 through 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 sampling frame and design stratum (Section 3.1)
- residential status adjustment for household eligibility (Section 3.2)
- adjustment for nonresponse to the CHIS household screener (Section 3.3)
- adjustment for reconsent of Asian language interviews (Section 3.4)
- adjustment for undersampling of listed 65+ sample (Section 3.5)

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 adult nonresponse (Section 4.2)
- adjustment to align the weight sums to population counts by telephone-usage status (Section 4.3)
- adjustment to combine cell phone and landline samples (Section 4.4)
- adjustment to combine the cell/landline sample with a supplemental sample in Imperial County (Section 4.5)
- 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) were 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)
- adjustment to account for differing probabilities of selection based on the number of adults, parents and children in the household as well as the age of the children (Section 5.2)
- adjustment for child nonresponse (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 frames (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 5.8)

The teen analysis weights (Chapter 6) were 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 frames (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 6.7)

Sections 4.6, 5.6, and 6.6 describe the final analysis weight which sometimes included constraints on outlier weights (Section 2.4).

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 2017-2018 Methodology Series: Report 3 - Data Processing Procedures*).

Chapter 9 contains a discussion on variance estimation for CHIS 2017-2018. 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 2017-2018, a model-based approach was implemented with the SUDAAN® WTADJUST procedure (RTI, 2012).

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 — 44.2% of California households were estimated in 2017 to have a landline telephone number in addition to at least one cell phone (National Center for Health Statistics, 2019). Conversely, only 4.3% of California households in 2017 were estimated to be landline only (National Center for Health Statistics, 2019). 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 $\hat{Y}^A$ to be the population parameter for the landline household population in California and $\hat{Y}^B$ to be the population parameters for the cellular household population in California for a statistic Y (e.g., total, mean). A CHIS estimate from the landline sampling frame would produce $\hat{Y}^A = \hat{Y}_a^A + \hat{Y}_{ab}^A$, where $\hat{Y}_a^A$ is the estimate from landline-only households and $\hat{Y}_{ab}^A$ is the frame-specific estimate for the dual users. Similarly, the associated unadjusted estimate from the cellular frame is $\hat{Y}^B = \hat{Y}_b^B + \hat{Y}_{ab}^B$where $\hat{Y}_b^B$ is the estimate from cell-only households and $\hat{Y}_{ab}^B$is the associated estimate for dual users from the cellular frame. The combined estimate, accounting for the overlap in the frames, should be estimated as

$$\hat{Y} = \hat{Y}_a^A + \lambda\hat{Y}_{ab}^A + (1-\lambda)\hat{Y}_{ab}^B + \hat{Y}_b^B \qquad (2.1)$$

where λ $(0 \leq \lambda \leq 1)$ is the composite factor 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. 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 4.4, 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 2017, an addressed-based sample (ABS) was chosen for a supplemental sample within a targeted area of Imperial County (see *CHIS 2017-2018 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 4.5, the child adjustment in Section 5.6, and the teen adjustment in Section 6.5.

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 by stratum were compared before and after each 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, unequal weighting effect) 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 (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 2017-2018 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 2017-2018 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 2017-2018 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 chapters of this report.

Specifically, we define the initial base weights by sampling frame in Section 3.1 that accounts for sampling at the household level. Section 3.2 contains an adjusted for unknown residential status. Weights for those with unknown residential status were then set to zero. Next, we applied an adjustment for household-level nonresponse defined as households without a completed screener (Section 3.3) followed by an Asian language reconsent adjustment (Section 3.4). The final adjustment in the household weighting was to account for the under sampling of listed 65+ households (Section 3.5). The final household weight is defined in Section 3.6.

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.

3.1.1 Landline Sample Base Weights

The base weight for the landline telephone sample, including leveraged listed Korean and Vietnamese surname samples, is computed as the inverse of the probability of selection of the telephone number. In CHIS 2017-2018, telephone numbers were drawn from a landline RDD frame, and a listed frame consisting of Korean and/or Vietnamese (V/K) surnames. The base weights reflect the multiple probabilities of selection of telephone numbers between the RDD and list frames.

The landline sampling was done in three steps within each stratum. The first step was to draw a SRS from among all RDD numbers. The second step was to draw additional samples of [a] numbers from high density V/K exchanges and [b] numbers from V/K surname lists. Note that the two supplemental frames overlap with each other and are subsets of the RDD frame. So, the RDD frame within each stratum can be divided into four partitions.

1. Numbers that are in high density V/K exchange and are listed V/K
2. Numbers that are in high density V/K exchange and are not listed V/K
3. Numbers that are not in high density V/K exchange and are listed V/K
4. Numbers that are not in high density V/K exchange and are not listed V/K

Table 3-1. RDD Landline Frame Partitions

Partition	Listed V/K	High density V/K
1	Yes	Yes
2	Yes	No
3	No	Yes
4	No	No

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

In effect, one sample is drawn from each of the four partitions. The only difference between the samples is that they are drawn from the high density V/K and the listed V/K partitions (1, 2 and 3) at higher rates.

As has been noted in past CHIS weighting reports, each of these are random samples and can be viewed as one overall sample. Overall then, the telephonic universe can be broken into not just the principal and substrata by which CHIS is sampled (44 main strata plus 8 Los Angeles substrata and 6 San Diego substrata) but as well by two demarcations of sample, sample that is listed V/K or not and sample that is in high density V/K exchanges or not.

Creating weights then requires the classification of every telephone number sampled by which partition it is in irrespective of how it was sampled. The resulting weights are composite weights derived by averaging the landline and list samples using a composite factor proportional to the sample sizes. Thus, this base weight produces an unbiased estimate in the traditional design-based framework.

The total universe of telephone numbers in the landline frame are computed using a standard RDD list-assisted approach of all 100 banks in the state of California with at least three directory-listed telephone numbers. The frame size is defined simply as the total universe of numbers available at the time of initial generation.

Overall then, the landline base weight adjustment is a simple sample fraction:

$$BW_{sp(l)i} = \frac{N_{sp(l)}}{n_{sp(l)}} \tag{3.1}$$

where s is the stratum and $p(l)$ is the landline frame partition within stratum. $N_{sp(l)}$ is the size of the RDD landline sample frame in partition $p(l)$ of stratum s and $n_{sp(l)}$ is the total amount of landline sample drawn from partition $p(l)$ of stratum s, including from the RDD sample, the additional high density V/K and the listed V/K sample.

3.1.2 Cell Phone Sample Base Weight

The cell phone sample was a stratified random sample drawn from 1,000-series blocks dedicated to wireless service. The sampling strata were defined by the cellular rate center of telephone numbers assigned to wireless service and pre-assigned FIPS county code.

The cell sample base weight is computed in the same way as the landline base weight described in Section 3.1.1. Note that the stratum definition for the cell phone sample is different from that of the landline sample in that they do not match the same geographic areas as in the landline sample for most strata. They do however include separate values for Los Angeles and San Diego substrata, and geographic areas with high concentrations of Koreans and Vietnamese were similarly oversampled on cell phones, and as well, listed surname sample of Koreans and Vietnamese were similarly leveraged in the cell phone frame.

As such the base weight adjustment is a simple sample fraction as with the landlines (see Equation (3.1)). However, there is one additional complication related to a new feature of the 2017-2018 CHIS which is the inclusion of a listed out of area cell sample of telephone numbers that are assigned to a rate center outside of California, but with an address appended that indicates that the owner of that cell phone does in fact live in California. These numbers were drawn from a consumer database of cell phone numbers that have an appended ZIP code. These numbers are associated with a geography outside of California but have a billing ZIP code in California. This sample helps capture cell phone users who currently live in California, but whose cell phone number is associated with a different state and would otherwise be excluded from an RDD sample of California.

We deal with this additional sample by defining address listed sample as the sum of all address listed cell sample in the CA universe plus all non-CA (area code) cell numbers available that have a CA listed address. In short, this creates a partition of the cell phone RDD frame into 8 potential cells detailed in the following table. Note that cell phone partitions 2 and 4 are empty because listed V/K numbers are a subset of all listed numbers, so there can be no numbers that are listed V/K, but not listed. Again the total cell sample universe is the sum of all CA rate center 1000 bank telephones plus all out-of-area cell phones; and each cell in the overall table is a result of subtracting total out of area listed universe counts by the total sample frame counts.

Table 3-2. RDD Cellular Frame Partitions

Partition	Listed V/K	High density V/K	Listed
1	Yes	Yes	Yes
2 - Empty	Yes	Yes	No
3	Yes	No	Yes
4 - Empty	Yes	No	No
5	No	Yes	Yes
6	No	Yes	No
7	No	No	Yes
8	No	No	No

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

The cellular base weight adjustment is a simple sample fraction:

$$BW_{sp(c)i} = \frac{N_{sp(c)}}{n_{sp(c)}} \qquad (3.2)$$

where s is the stratum and $p(c)$ is the cell frame partition within stratum. $N_{sp(c)}$ is the size of the cell sample frame in stratum s and $n_{sp(c)}$ is the total amount of cell sample drawn from stratum s, including the RDD sample, the additional high density V/K sample, the listed V/K sample and the address listed sample.

3.1.3 Address Based Sample Base Weight

For 2017 data collection, an ABS sample was drawn from a frame of mailing addresses in northern Imperial County. The base weight for each piece of ABS sample i is computed as:

$$BW_{NICi} = \frac{N_{NIC}}{n_{NIC}} \tag{3.3}$$

where N_{NIC} is the total number of ABS records in northern Imperial County and n_{NIC} is the amount of sample drawn from northern Imperial County. For the 2017 sample, $N_{NIC} = 10{,}746$ and $n_{NIC} = 5{,}252$.

3.1.4 AIAN Oversample Base Weight

There were two AIAN oversamples in the 2018 CHIS sample. The first of these consisted of listed landline and cell samples pulled from nine strata defined by three regions (Central, North and South) crossed by three AIAN densities that were based on AIAN incidences at the census block group level.

The base weight for each AIAN oversample case i in stratum h is computed as:

$$BW_{AIANhi} = \frac{N_{AIANh}}{n_{AIANh}} \tag{3.4}$$

where N_{AIANh} is the listed frame size in stratum h and n_{AIANh} is the amount of sample drawn from stratum h. Base weight adjustments were computed for the landline and cell listed frames separately.

This sample was combined with the base sample using a compositing factor to adjust for the proportion of sample used from each frame. This adjustment, λ_{AIAN} , is computed at the region level as follows.

$$\lambda_{AIAN} = \begin{cases} n_{LISTED,AIAN}/\left(n_{LISTED,AIAN} + n_{OS,AIAN}\right), & \text{for base sample listed AIAN cases} \\ n_{OS,AIAN}/\left(n_{LISTED,AIAN} + n_{OS,AIAN}\right), & \text{for AIAN oversample cases} \\ 1, & \text{all other sample} \end{cases} \tag{3.5}$$

where $n_{LISTED,AIAN}$ is the amount of main RDD landline sample that is flagged as listed and $n_{OS,AIAN}$ is the amount of AIAN listed oversample.

The second AIAN oversample consisted of 1,378 phone numbers randomly drawn from RDD landline telephone exchanges that mapped onto Indian reservations. The base weight for each piece of Indian Reservation (IR) sample i is computed as:

$$BW_{IRi} = \frac{N_{IR}}{n_{IR}} = \frac{9{,}700}{1{,}378} \cong 7.0392 \qquad (3.6)$$

where N_{IR} is the total amount of RDD landline numbers in telephone exchanges that map to the Indian reservations and n_{IR} is the amount of oversample drawn.

The second AIAN oversample was combined with the base landline sample using a compositing factor to adjust for the proportion of sample used from each frame. This adjustment, λ_{IR} , is computed as follows.

$$\lambda_{IR} = \begin{cases} n_{RDD,IR}/\left(n_{RDD,IR} + n_{OS,IR}\right), & \text{for base sample RDD landline cases in IR exchanges} \\ n_{OS,IR}/\left(n_{RDD,IR} + n_{OS,IR}\right), & \text{for IR oversample landline cases} \\ 1, & \text{all other sample} \end{cases} \qquad (3.7)$$

where $n_{RDD,IR}$ is the amount of RDD landline sample from the main sample that is from the IR exchanges and $n_{OS,IR}$ is the amount of IR oversample.

3.1.5 San Francisco Oversample Base Weight

For the San Francisco oversample, the landline and cell RDD frames were each divided into five strata. The base weight for each piece of San Francisco oversample is computed as the size of the strata divided by the size of the sample.

$$BW_{SFhi} = \frac{N_{SFh}}{n_{SFh}} \qquad (3.8)$$

where N_{SFh} is the frame size of stratum h and n_{SFh} is the amount of oversample drawn in stratum h. This sample was combined with the main landline and cell San Francisco samples using the following compositing factor:

$$\lambda_{SF} = \begin{cases} n_{RDD,SF}/\left(n_{RDD,SF} + n_{OS,SF}\right), & \text{for base sample RDD sases in San Francisco} \\ n_{OS,SF}/\left(n_{RDD,SF} + n_{OS,SF}\right), & \text{for oversample cases in San Francisco} \\ 1, & \text{all other sample} \end{cases} \qquad (3.9)$$

where $n_{RDD,SF}$ is the amount of RDD sample drawn from San Francisco and $n_{OS,SF}$ is the amount of San Francisco oversample drawn.

3.2 Residential Status Adjustment

Telephone numbers in the landline and cell phone samples with unknown residential status are those that could not be classified by residential status at the end of data collection despite being dialed many times. They are telephone numbers with only answering machine results or some combination of answering machine and ring no answer/busy results. Before adjusting the weights to account for telephone numbers with unknown residential status, the proportion of eligible residential telephone numbers among those numbers with unknown residential status was estimated. This was computed separately for the landline and cell phone samples.

The proportion of eligible residential telephone numbers (p_{res}) was computed following the CASRO recommendation as the proportion of the resolved or observed sample units that are residential. Since telephone numbers were sampled with different selection probabilities, the base-weighted number of telephone numbers rather than the (unweighted) number of cases was used to compute residential status.

We followed the method used in the 2013-2014 CHIS, namely we developed a weighting adjustment based on a combination of paradata in the sample. It has been found historically in past CHIS cycles that mail status, urban status[3], and answering machine status (none, residents, unknown) led to significantly different estimates, and thus served as effective variables by which to execute this adjustment. Tables 3-3, 3-4, and 3-5 summarize p_{res} for the landline, cell, and listed sample, respectively.

[3] Urban areas are defined as the center cites of Metropolitan Statistical Areas.

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

Mail status	Urban status	Voice mail status	p_{res17}	p_{res18}
Not mailable	Not urban	No voicemail	0.23	0.33
Not mailable	Not urban	Possible residential	0.77	0.88
Not mailable	Not urban	Unknown	0.64	0.75
Not mailable	Urban	No voicemail	0.12	0.22
Not mailable	Urban	Possible residential	0.86	0.84
Not mailable	Urban	Unknown	0.61	0.64
Mailable	Not urban	No voicemail	0.71	0.73
Mailable	Not urban	Possible residential	0.90	0.94
Mailable	Not urban	Unknown	0.86	0.90
Mailable	Urban	No voicemail	0.51	0.59
Mailable	Urban	Possible residential	0.95	0.94
Mailable	Urban	Unknown	0.86	0.86

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Table 3-4. Estimated residential proportion for the cell sample

Mail status	Urban status	Voice mail status	p_{res17}	p_{res18}
Not mailable	Not urban	No voicemail	0.35	0.27
Not mailable	Not urban	Possible residential	0.87	0.61
Not mailable	Not urban	Unknown	0.70	0.36
Not mailable	Urban	No voicemail	0.38	0.34
Not mailable	Urban	Possible residential	0.93	0.57
Not mailable	Urban	Unknown	0.80	0.37
Mailable	Not urban	No voicemail	0.51	0.38
Mailable	Not urban	Possible residential	0.93	0.66
Mailable	Not urban	Unknown	0.79	0.47
Mailable	Urban	No voicemail	0.51	0.42
Mailable	Urban	Possible residential	0.95	0.65
Mailable	Urban	Unknown	0.86	0.49

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Table 3-5. Estimated residential proportion for the listed sample

Voice mail status	p_{res17}	p_{res18}
No voicemail	0.42	0.41
Possible residential	0.95	0.68
Unknown	0.82	0.49

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Unlike 2013-2014, we were able to enact this procedure on the landline and cell RDD samples and the supplemental listed samples. This was not possible in the past because there was not mail status on cellphones, and the survey vendor did not attain sufficient information to make breaks by answering machine status. As was done in 2015-2016 CHIS, all cell phone numbers that were no answers on all attempts were classified as nonworking numbers.

The residential status adjusted weight, $HHA1W_i$, is

$$HHA1W_i = HHA1F_i \times BW_i \tag{3.10}$$

The residential status adjustment, $HHA1F_i$, was computed as follows:

$$HHA1F_i = \begin{cases} \left(\sum_{i \in RES} BW_i + \sum_{i \in UNK\_RES} p_{res} \times BW_i\right) / \sum_{i \in RES} BW_i, & \text{if } i \in \text{RES} \\ 0, & \text{if } i \in \text{UNK\_RES, NON\_RES} \\ 1, & \text{if } i \in \text{ABS sample} \end{cases} \tag{3.11}$$

where RES denotes telephone numbers identified as residential, UNK_RES denotes telephone numbers with unknown residential status, and NON_RES denotes non-residential telephone numbers. BW_i is the final base weight described in Section 3.1. p_{res} is the proportion of eligible residential telephone numbers.

3.3 Screener Nonresponse Adjustment

In this step, the household weights in the landline and cell phone samples were adjusted to account for households that did not complete the screener interview. The weights of nonresponding and ineligible households were distributed among eligible responding households. This weight, $HHA2W_i$, is computed as:

$$HHA2W_i = HHA2F_i \times HHA1W_i \tag{3.12}$$

where $HHA2F_i$ is the screener nonresponse adjustment factor. This adjustment was computed differently in 2017 and 2018, because of the large differences in eligibility rates between the main sample and the AIAN and San Francisco oversamples conducted in 2018. For the 2017 weighting, the screener nonresponse adjustment factor was computed as

$$HHA2F\_17_i = \begin{cases} \sum_{i \in SR,SNR} HHA1W_i \times \delta_i(c) / \sum_{i \in SR} HHA1W_i \times \delta_i(c), & \text{if } i \in \text{SR} \\ 0, & \text{if } i \in \text{SNR} \end{cases} \tag{3.13a}$$

where SR is the set of screener respondents and SNR is the set of screener nonrespondents. $\delta_i(c)$ defines the screener nonresponse adjustment groups based on sample frame, voice mail status and mailing status.

For the 2018 weighting, the screener nonresponse adjustment factor was computed as

$$HHA2F\_18_i = \begin{cases} \delta_i(c) \times \left(\sum_{i \in SRE} HHA1W_i + \sum_{i \in SNR} p_{elig} HHA1W_i\right) / \sum_{i \in SRE} HHA1W_i \times \delta_i(c), & \text{if } i \in \text{SRE} \\ 0, & \text{if } i \in \text{SNR, SRI} \end{cases} \quad (3.13b)$$

where SRE is the set of eligible screener respondents, SRI is the set of ineligible screener respondents, and SNR is the set of screener nonrespondents. p_{elig} is the proportion of screener responding households that were eligible and is computed within each screener nonresponse group.

The 2017 nonresponse adjustment groups were chosen based on a CART (Classification and Regression Trees) analysis.[4] For 2018, we used the same nonresponse adjustment groups that were used in 2017, but we broke out the AIAN and San Francisco oversamples into their own groups as outlined in the following table (Table 3-6).

Table 3-6. Screener Response Groups

2017 Screener Response Groups	2018 Screener Response Groups	
1 Landline, no residential VM, no address	1 Landline, no residential VM, no address	not AIAN or SF oversample
2 Landline, no residential VM, address	2 Landline, no residential VM, address	not AIAN or SF oversample
3 Landline, residential VM, no address	3 Landline, residential VM, no address	not AIAN or SF oversample
4 Landline, residential VM, address	4 Landline, residential VM, address	not AIAN or SF oversample
5 Cell, no residential VM, no address	5 Cell, no residential VM, no address	not AIAN or SF oversample
6 Cell, no residential VM, address	6 Cell, no residential VM, address	not AIAN or SF oversample
7 Cell, residential VM, no address	7 Cell, residential VM, no address	not AIAN or SF oversample
8 Cell, residential VM, address	8 Cell, residential VM, address	not AIAN or SF oversample
	9 AIAN oversample landline	
	10 AIAN oversample cell	
	11 SF oversample landline	
	12 SF oversample cell	

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

The screener nonresponse (SNR) adjustment is different for the ABS sample, and depends on the procedure used to obtain the telephone number for the sampled addresses and the different types of

[4] The CART analysis included five variables: sample frame; listed status; mail status; urban status and voicemail status. The three variables found to be significant – sample frame, mail status and voicemail status – were used to define the nonresponse groups.

nonresponse during this process. After selection, the sampled addresses were matched to telephone numbers using reverse telephone matching services. The remaining addresses without a matched telephone number were mailed a screening questionnaire asking for a telephone number associated with the address. All available telephone numbers, whether from the matching process or the mail screener, were dialed.

Nonresponse occurred in those households that did not return the mail screener, those households with telephone numbers that were not contacted, and those contacted households that refused the telephone interview.

Respondents were either those households that returned the mailed screener with a working phone number and completed the screener interview when contacted for the telephone interview, or that completed the screener interview when contacted through a matched telephone number.

The ABS screener nonresponse adjustment factor, $HHA2F_i$, was computed as

$$HHA2F_i = \begin{cases} \sum_{i \in A_{ER} A_{ENR} A_{IN}} HHA1W_i \times \delta_i(c) / \sum_{i \in A_{ER} A_{ENR}} HHA1W_i \times \delta_i(c), & \text{if } i \in \text{A\_ER in ABS frame} \\ 0, & \text{if } i \in \text{A\_ENR, A\_IN in ABS sample} \end{cases} \quad (3.14)$$

where A_ER, A_ENR, and A_IN are defined in Table 3-7. For the ABS sample, the screener nonresponse adjustment groups, $\delta_i(c)$, were based on the sample frame (landline or cell). The screener nonresponse adjustment factor for the ABS sample was incorporated into the weighting the same way it was incorporated for the telephone samples.

Table 3-7. ABS sample screener response groups

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

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

3.4 Asian Language Reconsent Adjustment

Some Asian language interviews conducted during 2018 data collection were not read the entire informed consent script (for more details, see Section 7.7 in *CHIS 2017-2018 Methodology Series: Report 2 – Data Collection Methods)*. These respondents were recontacted and read the entire informed consent script. A reconsent adjustment was made to account for cases that could not be recontacted and were dropped from the data. The reconsent adjustment weight, $HHA2AW_i$, was computed as follows:

$$HHA2AW_i = HHA2AF_i \times HHA2W_i \qquad (3.15)$$

where $HHA2AF_i$ is the Asian non-consent adjustment factor computed as

$$HHA2AF_i = \begin{cases} \sum_{i \in AR,ANR} HHA2W_i / \sum_{i \in AR} HHA2W_i\,, & i \in AR \\ 0, & i \in ANR \\ 1, & i \in NA \end{cases} \qquad (3.16)$$

where AR is the set of Asian reconsent cases that were successfully reconsented, ANR is the set of Asian reconsent cases for which we were not able to get reconsent, and NA are all the remaining cases for which no reconsent was necessary.

3.5 Listed 65+ Adjustment

A random 75% of all RDD landline sample and a random 75% of all RDD cell sample for which age information indicated there were only persons age 65 or older in the household were discarded in an effort to attain a younger overall sample. We adjust the weights to account for this undersampling of listed 65+ phone numbers and compute a listed 65+ adjusted weight, $HHA3W_i$ as follows:

$$HHA3W_i = HHA3F_i \times HHA2AW_i \tag{3.17}$$

where $HHA3F_i$ is the listed 65+ adjustment factor computed as:

$$HHA3F_i = \begin{cases} 4, & \text{if } i \in \text{DL65} \\ 1, & \text{if } i \in \text{NL65} \end{cases} \tag{3.18}$$

where DL65 is listed 65+ sample that was dialed and NL65 is sample that was not listed 65+ and was dialed.

3.6 One-Year Household Weight

The final one-year household weight is a product of the base weight and the four adjustment factors:

$$HHW_i = BW_i \times HHA1F_i \times HHA2F_i \times HHA2AF_i \times HHA3F_i = HHA3W_i \tag{3.19}$$

4. ADULT WEIGHTING

A final weight was created for each adult extended interview. Below, we detail the approach used to calculate an analysis weight for adults. Specifically, we define the initial base weights for the randomly selected adult within the household in Section 4.1. Nonresponse to the adult interview request is addressed next (Section 4.2), followed by calibration to phone totals (Section 4.3) and compositing of the landline and cell phone frames (Section 4.4). In Section 4.5, we describe a composite factor needed to combine the ABS and landline/cell samples within Imperial County. The weights for the entire sample are trimmed (Section 4.6) and then calibrated to estimated population projections (Section 4.7). The final adult analysis weight is summarized in Section 4.8. Statistics for the adjustments and the final adult weights are provided in Appendix B.

4.1 Number of Adults Adjustment

The first adjustment in the adult weighting adjusts for the number of adults in the household. For landline telephone numbers and those sampled from the ABS frame, one adult was selected with equal probability from all those residing in the household whereas on cell phones the respondent was assumed to be the sole owner and was selected for the interview. Thus, the number of adults adjustment is simply equal to the number of adults in the household for the landline and ABS samples. Since we consider the cell phone an unshared personal device, we assign a number of adults adjustment to 1.0 for cell cases.

As a result, the number of adults base weight, $ADA0W_i$, is defined as the product of the total household weight, HHW_i, and the number of adults adjustment factor, $ADA0F_i$:

$$ADA0W_i = ADA0F_i \times HHW_i \tag{4.1}$$

The number of adults adjustment factor, $ADA0F_i$ is

$$ADA0F_i = \begin{cases} AD_i, & \text{if } i \in \text{LL or ABS sample}, i = 1,2,3 \\ 1, & \text{if } i \in \text{Cell sample} \end{cases} \tag{4.2}$$

where AD_i is the number of adults in the household for respondent i. Consistent with past renditions of CHIS, values greater than three were truncated to an upper bound of three to limit the variation in the weight.

4.2 Adult Nonresponse Adjustment

Some households completed the screener interview, but the sampled adult did not complete the extended adult interview. To account for sampled adults who did not complete the extended interview, we include an adjustment for extended interview nonresponse. This was accomplished via a standard weighting class correction by specified groups.

A CART model was run to determine which variables best predicted adult response. The variables included in the model were those that were used in past waves of CHIS (sex of respondent, child first interview, language and adult screener respondent), plus we included sample frame (landline vs. cell). The variables that the CART model identified as significant were language, adult screener respondent and sample frame (see Table 4-1). Cells were collapsed due to small sample size (n < 25) when necessary.

Table 4-1. Variables used to create nonresponse adjustment cells for adult weights

Variable	Levels
Sample frame	1=landline, 2=cell
Language	1=English, 2=non-English
Adult screener respondent	1=sampled adult was screener R, 2=not

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

The adult nonresponse adjustment weight, $ADA1W_i$ is the product of the number of adults adjustment weight, $ADA0W_i$, and the adult nonresponse adjustment factor, $ADA1F_i$.

$$ADA1W_i = ADA1F_i \times ADA0W_i \tag{4.3}$$

The adjustment factor was a simple cell-based response propensity:

$$ADA1F_i = \begin{cases} \sum_{i \in ER,IN,UNK} ADA0W_i \times \delta_i(c) \Big/ \sum_{i \in ER,IN} ADA0W_i \times \delta_i(c), & \text{if } i \in \text{ER} \\ 0, & \text{if } i \in \text{UNK, IN} \end{cases} \tag{4.4}$$

where ER are eligible respondents who completed the extended interview, IN are ineligible cases identified during screening, and UNK = cases with unknown eligibility. $\delta_i(c) = 1$ if the adult is in adjustment cell c and $\delta_i(c) = 0$ otherwise.

4.3 Phone Use Calibration

A phone use calibration corrects for disproportionality between the population and sample distributions of those who only own landlines or cell phones, or both. A calibration adjustment was used to align the weight sums for the three groups generated from the telephone samples: landline only (LLO), cell phone only (CPO), and dual use respondents (DU). ABS sample did not receive this adjustment and was set to one.

As in past rounds of CHIS, we use 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 (National Center for Health Statistics, 2017).

$$ADA2W_i = ADA2F_i \times ADA1W_i \qquad (4.5)$$

The phone use calibration adjustment, $ADA2F_i$ was calculated as:

$$ADA2F_{gi} = \begin{cases} {N_g}/{\sum_{i \in g} ADA1W_i}, & \text{if } i \in \text{LL or Cell frame} \\ 1, & \text{if } i \in \text{ABS frame} \end{cases} \qquad (4.6)$$

where g denotes landline only, dual users or cell phone only and N_g denotes the population totals based on NHIS estimates.

As this adjustment was made separately for the landline and cell samples, the sum of the weights for the dual users across both frames once they are combined is double the population total. The next composite adjustment addresses this by creating a composite factor for the dual users in the combined landline/cell sample.

4.4 Composite of Cell Phone and Landline Samples

The CHIS 2017-2018 set the number of cell phone interviews at 50% of all interviews, both in total and within stratum (though quotas were not “hard” within strata, efforts were made to meet this general target). To adjust for this setting, a composite factor adjusts for the proportion of sample used from each frame (see Burton and Harter, 2015).

The frame composite weight, $ADA3W_i$, was the product of the phone use calibrated weight, $ADA2W_i$, and the frame composite adjustment, $\lambda 1_i$.

$$ADA3W_i = \lambda 1_i \times ADA2W_i \tag{4.7}$$

This adjustment, $\lambda 1$, is assigned at the strata level and was computed as follows:

$$\lambda 1 = \begin{cases} n_{DLL}/(n_{DLL} + n_{DCP}), & \text{if dual users in LL frame} \\ n_{DCP}/(n_{DLL} + n_{DCP}), & \text{if dual users in Cell frame} \\ 1, & \text{if LLO and CPO from LL and Cell frames} \\ 1, & \text{if ABS frame} \end{cases} \tag{4.8}$$

where n_{DLL} is the number of dual users in the landline frame and n_{DCP} is the number of dual users in the cell frame.

4.5 North Imperial County Composite Weight

The 2017 data collection included an oversample of north Imperial County drawn from an address-based sample. The independent samples from the main CHIS 2017 and the northern Imperial County ABS oversample need to be combined statistically. A composite factor adjusts for the overlap of the sample frames in northern Imperial County.

The northern Imperial County composite weight, $ADA4W_i$, was the product of the frame composite weight and the northern Imperial County composite adjustment, $\lambda 2$.

$$ADA4W_i = ADA3W_i \times \lambda 2 \tag{4.9}$$

The composite factor, $\lambda 2$, was defined as:

$$\lambda 2 = \begin{cases} n_{RDD}/(n_{RDD} + n_{ABS}), & \text{for northern Imperial County in RDD frame} \\ n_{ABS}/(n_{RDD} + n_{ABS}), & \text{for northern Imperial County in ABS frame} \\ 1, & \text{for not northern Imperial County} \end{cases} \tag{4.10}$$

As with (4.8), the lambda value for this compositing step was calculated using the relative unweighted sample sizes.

4.6 Pre-Calibration Trimming

The adult weight to this point is a product of the base weight from section 3 and the adjustments noted from Sections 4.1 to 4.5. This resulting weight for 2017 data was trimmed at the 2nd and 98th percentiles within strata. A total of 863 weights were trimmed across the 21,153 cases. None of the 2018 weights were trimmed prior to calibration.

4.7 Calibration Adjustment to Department of Finance Projections

We calibrated the composite weights 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 documented in Chapter 7 of this report. The calibrated weight was calculated as:

$$ADA5W_i = ADA4W_i \times AA1_i \qquad (4.11)$$

where $AA1_i$ is the calibration adjustment from the WTADJUST procedure.

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).

4.8 Adult One-Year Analysis Weight

The resulting adult weights, $ADA5W_i$, is the final one-year adult weight. There was no trimming done after the WTADJUST procedure was run.

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 2017 Imperial County oversample (Section 5.6). These weights were then trimmed (Section 5.7) and calibrated to population projections (Section 5.8). The child one-year analysis weight is shown 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 principal groups: Child-first households occur where the screener respondent is not the selected adult participant but is the legal guardian of the adult participant's children. On the other hand, surveys not employing the child first method are those where either the screener respondent is the selected adult respondent, or the screener respondent is not the selected adult respondent, but the latter is available to complete the survey. While child selection occurs in the screener for the child-first households, selection of a child occurs in Section G of the adult questionnaire for the other child households. As such sampling only occurs for "child last" cases if the adult completes Section G. Notably, the child-first methodology can only occur on landlines and ABS sample.

To account for adults who did not complete Section G of the adult interview, the household final weight, HHW_i, was adjusted by a Section G nonresponse adjustment factor, $HHA1\_1F_i$. We refer to this adjusted weight as the Section G adjusted household weight, $HHA1\_1W_i$

$$HHA1\_1W_i = HHW_i \times HHA1\_1F_i \tag{5.1}$$

The Section G nonresponse adjustment factor is computed as follows:

$$HHA1\_1F_i = \begin{cases} \sum_{i \in NC1stGC, NC1GNC} HHW_i \delta_i(c) \Big/ \sum_{i \in NC1stGC} HHW_i \delta_i(c), & \text{if } i \in NC1stGC \\ 0, & \text{if } i \in NC1stGNC \\ 1, & \text{if } i \in C1st \end{cases} \tag{5.2}$$

where NC1stGC refers to households without the child first procedure and Section G was completed, NC1stGNC refers to households without the child first procedure, and Section G was not completed, and C1st refers to households with the child-first procedure. c refers to the Section G nonresponse adjustment cell and $\delta_i(c) = 1$ if the adult belongs to cell c and $\delta_i(c) = 0$ otherwise.

5.2 Base Weights

The child base weights are necessary to account for the disproportionate sampling of children by age group within household. Specifically, children ages 0-5 were given twice the likelihood of selection than children 6-11 by study design. If $n1$ is the number of children age 0-5 of the sampled adult in the household and $n2$ is the number of children 6-11 of the sampled adult in the household, then probability that a child is sampled, $CHA0_i$, is defined as:

$$CHA0_i = \begin{cases} (2 \times n1_i) \Big/ \left((2 \times n1_i) + n2_i\right), & \text{for children ages } 0-5 \text{ years} \\ n2_i \Big/ \left((2 \times n1_i) + n2_i\right), & \text{for children ages } 6-11 \text{ years} \end{cases} \tag{5.3}$$

The child base weight also needs to account for the different probability of child selection across households based on the number of adults and parents in the households. Households with two parents have twice the probability of selecting a parent than households with only one parent (and other adults in the household). If we let P_i be the number of parents in household i, and AD_i the number of the adults in the household (capped at 3), then the resulting child-level base weight is defined as :

$$CHW0_i = HHA1\_1W_i \Big/ CHA0_i \times \left(P_i / AD_i\right) \tag{5.4}$$

5.3 Child Nonresponse Adjustment

Similar to CHIS weighting prior to 2015, we calculate a child nonresponse adjustment in the same manner as the adult nonresponse adjustment described in Section 4.2. The adjustment cells are defined by sex within sampling stratum. Small cells were collapsed cells to increase the number of respondents in each cell.

$$CHA1W_i = CHA1F_i \times CHW0_i \tag{5.5}$$

The adjustment factor, $CHA1F_i$, is:

$$CHA1F_i = \begin{cases} \sum_{i \in CHR,CHNR} CHW0_i \times \delta_i(c) \Big/ \sum_{i \in CHR} CHW0_i \times \delta_i(c), & \text{if } i \in CHR \\ 0, & \text{if } i \in CHNR \end{cases} \tag{5.6}$$

where CHR are child-interview respondents and CHNR are child interview non-respondents. We define c as the child nonresponse adjustment cell defined using sex and sampling stratum. $\delta_i(c) = 1$ if the case is in the adjustment cell and $\delta_i(c) = 0$ otherwise.

5.4 Phone Use Calibration

The child weight employs the same phone use calibration as the adult weight. Again, this calibration adjustment was implemented to align the weight sums for the three groups generated from the telephone samples: landline only (LLO), cell phone only (CPO), and dual use landline sample (DU). ABS sample did not receive this adjustment and was set to one.

$$CHA2W_i = CHA2F_i \times CHA1W_i \tag{5.7}$$

The phone use calibration adjustment, $CHA2F_i$, was calculated as:

$$CHA2F_{gi} = \begin{cases} N_g \Big/ \sum_{i \in g} CHA1W_i, & \text{if } i \in \text{LL or Cell frame} \\ 1, & \text{if } i \in \text{ABS frame} \end{cases} \tag{5.8}$$

where g denotes landline only, dual users or cell phone only and N_g denotes population totals based on the NHIS estimates similar to those used in Section 4.3. For the child calibration adjustment, NHIS estimates for children's phone use were used.[5]

As this adjustment is made separately for the landline and cell samples, the sum of the weights for the dual users across both frames once they are combined is double the population total. The next

[5] NHIS provides phone use estimates for 0-17 and does not break them out for children and teen separately. We used NHIS 0-17 estimates for both the child and teen calibrations.

composite adjustment will address this by creating a composite factor for the dual users in the combined landline/cell sample.

5.5 Composite of Cell Phone and Landline Samples

The child weight utilizes the same frame composite as the adult weight, again, a composite factor that adjusts for the proportion of sample used from each frame (see Burton and Harter, 2015).

The frame composite weight, $CHA3W_i$, was the product of the phone use calibrated weight, $CHA2W_i$, and a frame composite adjustment, $\lambda 1_i$. The same lambda values used for the adult weights (see Equation (4.8)) were applied to the child weights.

$$CHA3W_i = CHA2W_i \times \lambda 1_i \tag{5.9}$$

5.6 Composite of Cell/Landline and Northern Imperial County Samples

The 2017 child weight utilizes the same adjustment for North Imperial County as the 2017 adult weight. The North Imperial composite weight, $CHA4W_i$, was the product of the frame composite weight, $CHA3W_i$, and a composite factor, $\lambda 2_i$. The same lambda values used for the adult weights (see Equation (4.10)) were applied to the child weights.

$$CHA4W_i = CHA3W_i \times \lambda 2_i \tag{5.10}$$

5.7 Pre-Calibration Trimming

The child weight to this point is a product of the base weight from Chapter 3 and the adjustments noted from Section 5.1 to 5.6. There were 4 out of the 1,600 cases in the 2017 data that had outlier weights trimmed to reduce unequal weighting effects. Weights were trimmed in three counties – Contra Costa, Butte, and Imperial – where the 2017 unequal weighting effect was greater than 2.80. No child weights were trimmed prior to the 2018 calibration.

5.8 Calibration Adjustment to Department of Finance Projections

The child data was calibrated to target population parameters like the adult data. The calibrated weight was calculated as:

$$CHA5W_i = CHA4W_i \times AA2_i \tag{5.11}$$

where $AA2_i$ is the calibration adjustment from the WTADJUST procedure.

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).

5.9 Child One-Year Analysis Weight

The resulting child weight, $CHA5W_i$, is the final one-year child weight. There was no trimming done after the WTADJUST procedure was run.

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 weight 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 2017 Imperial County ABS areas is described in Section 6.5. Weight trimming is described in Section 6.6. We constructed a calibration adjustment to population projections (Section 6.7). Statistics for the adjustments and the final teen weights are provided in Appendix B.

6.1 Base Weights

As in the child weighting, the initial weights for the adolescents incorporate the probability of sampling the adult and the probability of sampling an adolescent among all adolescents associated with the sampled adult. The initial weight, $TN0W_i$, is computed as

$$TN0W_i = HHW_i \times {TCNT_i}\Big/\left({p_i}/{AD_i}\right) \qquad (6.1)$$

where p_i is the number of parents in household i, AD_i is the number of adults in the household (capped at 3), and $TCNT_i$ is the number of eligible teens of the sampled parent. HHW_i is the household weight defined in Section 3.6.

6.2 Adjustment for Teen Nonresponse

Nonresponse can occur due to a refusal of the parent to grant permission to interview the adolescent, as well as refusal from the adolescent. Historically these two types of nonresponse are combined due to small sample sizes. Thus this nonresponse adjustment was based on sample frame and stratum:

$$TNA1W_i = TNA1F_i \times TN0W_i \tag{6.2}$$

The adjustment factor, $TNA1F_i$, is:

$$TNA1F_i = \begin{cases} \sum_{i \in TNR,TNNR} TN0W_i \times \delta_i(c) \Big/ \sum_{i \in TNR} TN0W_i \times \delta_i(c), & \text{if } i \in TNR \\ 0, & \text{if } i \in TNNR \end{cases} \tag{6.3}$$

where TNR are teen-interview respondents and TNNR are teen interview non-respondents. We define *c* as the adolescent nonresponse adjustment cell defined using stratum. $\delta_i(c) = 1$ if the case is in the adjustment cell and $\delta_i(c) = 0$ otherwise. Due to the small sample size, the nonresponse adjustment cells are defined by sample stratum.

6.3 Calibration Adjustment to NHIS

The adolescent weight employs the same phone use calibration as the adult and child weights. Again this calibration adjustment is implemented to align the weight sums for the three groups generated from the telephone samples: landline only (LLO), cell phone only (CPO), and dual use sample (DU). ABS sample does receive this adjustment and was set to one.

$$TNA2W_i = TNA2F_i \times TNA1W_i \tag{6.4}$$

The phone use calibration adjustment, $TNA2F_{gi}$, is calculated as:

$$TNA2F_{gi} = \begin{cases} N_g \Big/ \sum_{i \in g} TNA1W_i, & \text{for } i \in \text{LL or Cell frame} \\ 1, & \text{for } i \in \text{ABS frame} \end{cases} \tag{6.5}$$

where g denotes landline only, dual users or cell phone only, and N_g denotes population totals based on the same NHIS estimates used in Section 5.4.

6.4 Composite of Cell Phone and Landline Samples

The adolescent weight utilizes the same frame calibration as the adult and child weights, again, a composite factor that adjusts for the proportion of sample used from each frame (see Burton and Harter, 2015).

The frame composite weight, $TNA3W_i$, was the product of the phone use calibrated weight, $TNA2W_i$, and a frame composite adjustment, $\lambda1_i$. The same lambda values used for the adult weights and child weights (see Equation (4.8)) were applied to the adolescent weights.

$$TNA3W_i = TNA2W_i \times \lambda1_i \tag{6.6}$$

6.5 Composite of Cell/Landline and Northern Imperial County Samples

The 2017 adolescent weight utilizes the same adjustment for northern Imperial County as the 2017 adult weight. The northern Imperial County composite weight, $TNA4W_i$, is the product of the frame composite weight, $TNA3W_i$, and a composite factor, $\lambda2_i$. The same lambda values used for the adult weights and child weights (see Equation (4.10)) were applied to the adolescent weights.

$$TNA4W_i = TNA3W_i \times \lambda2_i \tag{6.7}$$

6.6 Pre-calibration Trimming

The teen weight to this point is a product of the base weight from Chapter 3 and the adjustments noted from Section 6.1 to 6.5. There was 1 case out of the 448 teen completes from CHIS 2017 that had its weight trimmed because the untrimmed weight resulted in an unequal weighting effect of greater than 3.0. The trimmed case was in Imperial County. There were no cases trimmed for the 2018 teen weights.

6.7 Calibration Adjustment to Department of Finance Projections

The teen data was calibrated to target population parameters like the adult data. The calibrated weight was calculated as:

$$TNA5W_i = TNA4W_i \times AA3_i \tag{6.8}$$

where $AA3_i$ is the calibration adjustment from the WTADJUST procedure.

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).

6.8 Teen One-Year Analysis Weight

The resulting weight, $TNA5W_i$, is the final one-year adolescent weight.

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 covariates than poststratification but excluding finer adjustments that could benefit the survey estimates.

Calibration using the WTADJUST procedure in SUDAAN (Section 2.2.2) 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).

Calibration adjustments were used twice in CHIS 2017-2018. 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, 5.5, and 6.4, 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, 5.8, and 6.7, 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.

We ran 11 different calibrations to align weight sums to population estimates. We ran an untrimmed calibration along with calibrations that trimmed the weight at 1%, 2%,…,10%. We computed mean squared errors on a series of variables to decide on a final trimming.[6] There was no one trimming that resulted in a minimum mean squared error across all of the variables and differences among the trimmings were very subtle. We used the 1% trim as it minimized the MSE for the majority of the variables used in the analysis.

7.2 Calibration Model Dimensions

The 15 weight calibration dimensions used in CHIS 2017-2018 are shown in Table 7-1. These dimensions follow those specified in prior years of the study to maximize continuity. Specifically, Dimensions 1-8 and 11 involve 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 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. Dimension 9 includes education of the responding adult crossed with region and Dimension 10 includes number of adults in the household crossed by primary age crossed by region. Dimension 12 interacts age groups (3) with household phone-usage status (landline only, cell phone only, and dual user). Dimension 13 is complex and is described below. Dimension 14 involves the calculation of counts to incorporate the targeted area within Imperial County for CHIS 2017. Dimension 15 involves the oversample in San Francisco for CHIS 2018.

Levels within the dimensions were collapsed for situations where there were fewer than 50 respondents in a cell. Table 7.1 shows the 15 calibration dimensions along with the total number of categories for each. The last column of the table shows the number of categories that were used in the calibration after collapsing. Table 7.2 shows the definition of all the variables that were used to create the 15 dimensions.

---

[6] The variables used in the trimming analysis were DISTRESS, AB1, ASTCUR (adult), AB22, AH16, AH22, AI8, CA6, ASTCUR (child), TB1, and ASTCUR (teen).

Table 7-1. Dimensions used in Weight Calibration

Dimension	Variables (categories)	Total categories[1]	Categories after collapsing for 2017 and/or 2018 data	Categories after collapsing for combined 2017-2018 data
1	Region (7) by primary age 1 (3) by sex (2)	42	30	35
2	Region (7) by secondary age (9)	63	58	63
3	Detailed age (13) by sex (2)	26	26	26
4	Geography (14) by primary age 1 (3) plus remainder (1)	43	19	28
5	Primary age 2 (2) by race/ethnicity (7) by region (7)	98	34	60
6	Primary age 1 (3) by race/ethnicity (7) by sex (2)	42	26	30
7	Asian groups (6) by primary age 1 (3)	18	8	10
8	Stratum (44) by race (3) by primary age 2 (2)	264	88	133
9	Region (7) by education (6)	42	33	36
10	Region (7) by primary age 1 (3) adults in household (3)	63	37	44
11	Stratum (44) by primary age 1 (3)	132	55	73
12	Phone use (3) by primary age 1 (3)	9	8	8
13	Region (7) by phone proxy (16)	112	94	101
14 (2017)	Primary age in North Imperial county (3) plus remainder of Imperial county (1) plus remainder of state (1)	5	4	4
15 (2018)	Race by sex in San Francisco (5) plus remainder of state (1)	6	6	6

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

[1] The total number of categories for each dimension is simply the product of the individual variables used to create the dimension, plus any remainder categories (dimensions 4 and 14).

Table 7-2 details the variables used in the creation of the 15 calibration dimensions. The number of categories is listed in parenthesis followed by a list of the dimensions that use the variable.

Table 7-2. Detailed variable definitions used in calibration dimensions

Variable	Dimensions	Categories	
Region (7)	1,2,5,9,10,13	*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 counties	
		Greater Bay Area: Santa Clara, Alameda, Contra Costa, San Francisco, San Mateo, Sonoma, Solano, Marin, Napa counties	
		Sacramento Area: Sacramento, Placer, Yolo, El Dorado counties	
		San Joaquin Valley: Fresno, Kern, San Joaquin, Stanislaus, Tulare, Merced, Kings, Madera counties	
		Central Coast: Ventura, Santa Barbara, Santa Cruz, San Luis Obispo, Monterey, San Benito counties	
		Los Angeles: Los Angeles County	
		Other Southern California: San Diego, Orange, San Bernardino, Riverside, Imperial counties	
Primary age 1 (3)	1,4,6,10,11,12,14 (2017)	0-17 years	
		18-64 years	
		65+ years	
Sex (2)	1,6,13	Male	
		Female	
Secondary age (9)	2	0-5 years	30-39 years
		6-11 years	40-49 years
		12-17 years	50-64 years
		18-24 years	65+ years
		25-29 years	
Detailed age (13)	3	0-3 years	31-37 years
		4-7 years	38-45 years
		8-11 years	46-53 years
		12-14 years	54-64 years
		15-17 years	65-77 years
		18-24 years	78+ years
		25-30 years	

(continued)

Table 7-2. Detailed variable definitions used in calibration dimensions (continued)

Variable	Dimensions	Categories
Geography (14)	4	Los Angeles County – Antelope Valley
		Los Angeles County – San Fernando Valley
		Los Angeles County – San Gabriel Valley
		Los Angeles County – Metro
		Los Angeles County – West
		Los Angeles County – South
		Los Angeles County – East
		Los Angeles County – South Bay
		San Diego County – North Coastal
		San Diego County – North Central
		San Diego County – Central
		San Diego County – South
		San Diego County – East
		San Diego County – North Inland
Primary age 2 (2)	5,8	0-17 years
		18+ years
Race/ethnicity (7)	5,6	Latino
		White, not Latino
		Black, not Latino
		American Indian, not Latino
		Asian, not Latino
		Native Hawaiian, not Latino
		Two or more races, not Latino
Asian groups (6)	7	Not Latino Chinese
		Not Latino Korean
		Not Latino Filipino
		Not Latino Vietnamese
		Not Latino Japanese
		Not Latino other Asian
Stratum (44)	8,11	Refer to Table 1-1 for strata definitions
Race (3)	8	Latino
		Not Latino, White
		Not Latino, other race
Education (6)	9	Under 18 and parent less than HS graduate
		Under 18 and parent HS graduate
		Under 18 and parent some college+
		18+, less than HS graduate
		18+, HS graduate
		18+, some college+

(continued)

Table 7-2. Detailed variable definitions used in calibration dimensions (continued)

Variable	Dimensions	Categories
Number of adults in household (3)	10	One adult
		Two adults
		Three or more adults
Phone use (3)	12	Landline only
		Cell phone only
		Dual use
Phone proxy (16)	13	Home owner, 0-17, 0-1 adult
		Renter, 0-17, 0-1 adult
		Home owner, 0-17, 2+ adults
		Renter, 0-17, 2+ adults
		Home owner, 18-30, HS grad or less
		Home owner, 31-64, HS grad or less
		Home owner, 65+, HS grad or less
		Home owner, 18-30, greater than HS grad
		Home owner, 31-64, greater than HS grad
		Home owner, 65+, greater than HS grad
		Renter, 18-34, up to HS grad
		Renter, 35+, up to HS grad, 0-1 adult
		Renter, 35+, up to HS grad, 2+ adults
		Renter, 18-34, greater than HS grad
		Renter, 35+, greater than HS grad, 0-1 adult
		Renter, 35+, greater than HS grad, 2+ adults
Race by sex in San Francisco (5)	15 (2018)	San Francisco, Latino, Male
		San Francisco, Black, not Latino, Male
		San Francisco, Chinese, not Latino, Male
		San Francisco, Black, not Latino, Female
		Rest of San Francisco
		Rest of California

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Dimension 13 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 2.6% of households in California[7], 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.

[7] https://www.cdc.gov/nchs/data/nhis/earlyrelease/Wireless_state_201712.pdf

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 7-3) were provided for 2017 and 2018 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*.

Table 7-3. Definition of counts available in 2017 and 2018 California DOF population files

Category	Levels
County (58)	Alameda, Alpine, …, Yolo, Yuba
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 (12)	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 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: 2017-2018 California Department of Finance projections.

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, 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 three 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 7-4.
- 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 for producing population distributions for Dimension 4 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.

Table 7-4. Definition of variables available on the 2010 Census Summary File

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

American Community Survey (ACS) public-use one-year micro data files (PUMS) were accessed for Dimensions 7, 9, 10, 13, 14, and 15. These data were used to estimate the proportions of the population by Asian groups, education, household tenure, and number of adults in the household within the seven California regions (Table 7-2). Additionally, these data were used to estimate population proportions by age in northern Imperial County and by sex and race in San Francisco. The 2016 ACS PUMS file was used for CHIS 2017 and 2018 one-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 to estimate telephone service type for CHIS 2017-2018. The estimates were required for calibration Dimension 12 and for combining landline and cell phone samples (see, for example, Section 4.4 related to adult weight adjustments) and are shown in Table 7-5.

Table 7-5. NHIS proportions of telephone service by person type and year

Person type	Telephone service	Percent
Adult	Landline only	4.9
	Dual use[1]	42.5
	Cell phone only	50.0
	No phone	2.6
Child and Teen	Landline only	2.0
	Dual use[1]	36.5
	Cell phone only	58.7
	No phone	2.8

Source: NCHS, National Health Interview Survey, 2012–2016; U.S. Census Bureau, American Community Survey, 2011–2015; and infoUSA.com consumer database, 2012–2016.

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

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 14 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 six levels for race in the DOF file are shown in Table 7-3 and the 18 age levels required for the calibration dimensions are shown in Table 7-6. 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”.

Table 7-6. Age levels used to summarize California DOF data file

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

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.
Note: DOF = Department of Finance.

Estimated Group Quarters. The estimated proportion of group quarters was estimated from the 2010 Census SF1. As shown in Table 7-4, 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 total set was defined as

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

where D_{1m}^{SF1} was the total population of 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 of these variables are shown in Table 7-4.

The second set of control totals were defined as

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

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

The third set of controls were calculated as

$$D_{3q}^{SF1.\overline{GQ}} = D_{3q}^{SF1} - D_{3q}^{SF1.GQ} \quad (7.3)$$

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

$$T_{d7}^{sf1.\overline{GQ}} = T_{mp}^{SF1} \times a_{mp} \quad (7.4)$$

where T_{mp}^{SF1} were the 2010 Census SF1 population counts within cross-classified groups defined in Table 7-4, a_{mp} was the adjustment applied based on raking the counts to the control totals, and *d7* 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

$$p_{d7}^{SF1.\overline{GQ}} = T_{d7}^{SF1.\overline{GQ}} \big/ T_{d7}^{SF1} \quad (7.5)$$

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,

$$T_{d6}^{DOF.\overline{GQ}} = p_{d7}^{SF1.\overline{GQ}} \times T_{d6}^{DOF} \quad (7.6)$$

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 2.4%.

7.4.2 Computing the Control Totals

Values calculated with (7.6) were tabulated across the estimation cells to form the non-group quarters control totals for calibration dimensions 1, 2, 3, 5, 6, 8 and 11. 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 2016 ACS PUMS data and applied to the adjusted DOF counts. ACS data were also used for Dimensions 9 (adult's education), 10 (number of adults in the household), 13 (non-telephone adjustment), 14 (2017 Imperial County adjustment) and 15 (2018 San Francisco 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 2017-2018.

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 corresponding screener variable (SC62_AGE…SCF2_AGE).

Table 8-1. Description of imputed variables by year

Variable Type	Variable Name	Variable Description	Response Variables
Geographic (reported)	SR_COUNTY_FIPS	County	AH42, KAH42, SAH42
	SRZIP	ZIP Code	AM7, KAM7, SAM7
	SRSTRATA	Stratum	AH42, KAH42, SAH42
	SR_LASPA	Los Angeles Service Planning Area (SPA)	AH42, KAH42, SAH42, AM7, KAM7, SAM7
	SR_HR	San Diego Health Service Region (HSR)	AH42, KAH42, SAH42, AM7, KAM7, SAM7
Household	SRTENR	Household tenure	AK25, KAK25
	HASCELL	Cell/Wireless telephone service	AM33, KAM33, AN10, CELL8, SINTRO_3A
	HASLANDLINE	Landline telephone service	AN6, AN7, CELL8, SINTRO_3A
	CALLINTENSITY	Phone Use Intensity	AM34, HASCELL, HASLANDLINE
	ELIG_KID_0_5	Number of interview-eligible kids ages 0-5	SC13A2_01 –SC13A2_20, SC15A_1 –SC15A_20, SC14A1, SC14A_01-SC14A_20, SC14B_01 –SC14B_20, ADULT_INDEX, TEEN_INDEX, CHILD_INDEX
	ELIG_KID_6_11	Number of interview-eligible kids ages 6-11	SC13A2_01 –SC13A2_20, SC15A_1 –SC15A_20, SC14A1, SC14A_01-SC14A_20, SC14B_01 –SC14B_20, SC14C_01-SC14C_20, ADULT_INDEX, TEEN_INDEX, CHILD_INDEX
	ELIG_TEEN	Number of interview-eligible teens	SC13A2_01 –SC13A2_20, SC15A_1 –SC15A_20, SC14A1, SC14A_01-SC14A_20, SC14B_01 –SC14B_20, SC14C_01-SC14C_20, ADULT_INDEX, TEEN_INDEX, CHILD_INDEX
	PARENT_CHILD_HH	Number of parents for the selected child	SC14A_01-SC14A_20, SCB_01-SC14B_20, SC14C_01-SC14C_20, PERSNUM_CHILD
	PARENT_TEEN_HH	Number of parents for the selected teen	SC14A_01-SC14A_20, SCB_01-SC14B_20, SC14C_01-SC14C_20, PERSNUM_TEEN

(continued)

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

Variable Type	Variable Name	Variable Description	Response Variables
Person	SRAGE	Age	AAGE, SC62_AGE…SCF2_AGE, CAGE, TAGE
	SRSEX	Sex	AA3, CA1, TA3
	SREDUC	Educational Attainment	AH47, KAH47
	SRH	Self-Reported Latino	AA4, CH1, TI1
	SRW	Self-Reported White	AA5A_A - AA5A_G, CH3_A - CH3_G, TI2_a – TI2_G
	SRAA	Self-Reported African American	AA5A_A - AA5A_G, CH3_A - CH3_G, TI2_a – TI2_G
	SRAS	Self-Reported Asian	AA5A_A - AA5A_G, CH3_A - CH3_G, TI2_a – TI2_G
	SRAI	Self-Reported American Indian/Alaska Native	AA5A_A - AA5A_G, CH3_A - CH3_G, TI2_a – TI2_G
	SRPI	Self-Reported Native Hawaiian and Other Pacific Islander	AA5A_A - AA5A_G, CH3_A - CH3_G, TI2_a – TI2_G
	SRO	Self-Reported Other	AA5A_A - AA5A_G, CH3_A - CH3_G, TI2_a – TI2_G
	SRCH	Self-Reported Chinese	AA5E_A – AA5E_G, CH7_A – CH7_G, TI2D_A – TI2D_G
	SRPH	Self-Reported Filipino	AA5E_A – AA5E_G, CH7_A – CH7_G, TI2D_A – TI2D_G
	SRKR	Self-Reported Korean	AA5E_A – AA5E_G, CH7_A – CH7_G, TI2D_A – TI2D_G
	SRJP	Self-Reported Japanese	AA5E_A – AA5E_G, CH7_A – CH7_G, TI2D_A – TI2D_G
	SRVT	Self-Reported Vietnamese	AA5E_A – AA5E_G, CH7_A – CH7_G, TI2D_A – TI2D_G
	SRASO	Self-Reported Other Asian	AA5E_A – AA5E_G, CH7_A – CH7_G, TI2D_A – TI2D_G
	OMBSRREO	OMB Race/ Ethnicity Group	SRH, SRO, SRW2, SRAA2, SRAS2, SRAI2, SRPI2
	OMBSRASO	OMB non-Latino Asian Group	SRH, SRAS, SRCH, SRPH, SRKR, SRJP, SRVT, SRASO

Source: UCLA Center for Health Policy Research, 2017-2018 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, 2017-2018 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.

- *External data source assignment*: 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
SC5A	Number of adults in the household
CALLINTENSITY	Self-reported phone intensity
CHLD_INDEX	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
POVERTY	Poverty status
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
TEEN_INDEX	Presence of teens in the household

Source: UCLA Center for Health Policy Research, 2017-2018 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 ZIP Code

Self-reported 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 combination of external data source assignment and 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 zip code by interview mode

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[2]	n	pct[2]	n	pct[2]
SRZIP (Self-reported ZIP code)						
Reported values	35,973	85.0	16,365	73.6	19,608	97.6
Imputed values	174	0.4	165	0.7	9	0.0
External data source assignment	6,183	14.6	5,701	25.6	482	2.4
Total	42,330	100.0	22,231	100.0	20,099	100.0

Source: UCLA Center for Health Policy Research, 2017-2018 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 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 external data source assignment. 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

			Interview Mode[1]			
	All Modes		Cell		Landline	
Variable and Source of Data[1]	n	pct[2]	n	pct[2]	n	pct[2]
SRSTRATA (Self-reported stratum)						
Reported values	41,072	97.0	21,379	96.2	19,693	98.0
Imputed values	94	0.2	86	0.4	8	0.0
External data source assignment	1,164	2.7	766	3.4	398	2.0
Total	42,330	100.0	22,231	100.0	20,099	100.0
SR_LASPA (Self-reported Los Angeles county service planning area)						
Reported values	42,061	99.4	22,035	99.1	20,026	99.6
Imputed values	20	0.0	19	0.1	1	0.0
External data source assignment	249	0.6	177	0.8	72	0.4
Total	42,330	100.0	22,231	100.0	20,099	100.0
SR_HR (Self-reported San Diego county health service region)						
Reported values	42,171	99.6	22,115	99.5	20,056	99.8
Imputed values	15	0.0	14	0.1	1	0.0
External data source assignment	144	0.3	102	0.5	42	0.2
Total	42,330	100.0	22,231	100.0	20,099	100.0

Source: UCLA Center for Health Policy Research, 2017-2018 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.

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 phone usage (Landline vs. Cell phone only). 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	41,786	98.7	21,893	98.5	19,893	99.0
Imputed values	544	1.3	338	1.5	206	1.0
Total	42,330	100.0	22,231	100.0	20,099	100.0

Source: UCLA Center for Health Policy Research, 2017-2018 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 initial mode sample variable (LL vs. Cell), SRTENR and OMBSRREO, while the imputation classes for CALLINTENSITY were created from SRAGE and SRTENR. 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	41,811	98.8	22,196	99.8	19,615	97.6
Imputed values	519	1.2	35	0.2	484	2.4
Total	42,330	100.0	22,231	100.0	20,099	100.0
HASLANDLINE (Presence of a landline phone)						
Reported values	41,811	98.8	22,196	99.8	19,615	97.6
Imputed values	519	1.2	35	0.2	484	2.4
Total	42,330	100.0	22,231	100.0	20,099	100.0
CALLINTENSITY (Self-reported phone intensity)						
Reported values	41,811	98.8	22,196	99.8	19,615	97.6
Imputed values	519	1.2	35	0.2	484	2.4
Total	42,330	100.0	22,231	100.0	20,099	100.0

Source: UCLA Center for Health Policy Research, 2017-2018 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

Variable and Source of Data	All Modes		Interview Mode[1] Cell		Interview Mode[1] Landline	
	n	pct[2]	n	pct[2]	n	pct[2]
ELIG_KID_0_5 (Self-reported number of eligible children age 0-5)						
Reported values	23,458	55.4	12,619	56.8	10,839	53.9
Logically imputed values	18,665	44.1	9,468	42.6	9,197	45.8
Hot-deck imputed values	207	0.5	144	0.6	63	0.3
Total	42,330	100.0	22,231	100.0	20,099	100.0
ELIG_KID_6_11 (Self-reported number of eligible children age 6-11)						
Reported values	23,461	55.4	12,615	56.7	10,846	54.0
Logically imputed values	18,665	44.1	9,468	42.6	9,197	45.8
Hot-deck imputed values	204	0.5	148	0.7	56	0.3
Total	42,330	100.0	22,231	100.0	20,099	100.0

Source: UCLA Center for Health Policy Research, 2017-2018 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 there were no missing values in this variable, we did not need to employ hot-deck imputation.

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. As there were no missing values in these variables, they were not imputed.

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. As with the previous CHIS cycles, data for adult respondents who answered "unknown" to the household income questions were left unchanged. There were no other missing value requiring imputation.

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 8-9). This section describes the imputation procedures used for each variable needed for weighting and their item nonresponse rates.

Table 8-9. 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	42,330	22,231	52.5	20,099	47.5
Child	3,186	2,054	64.5	1,132	35.5
Teen	880	418	47.5	462	52.5

Source: UCLA Center for Health Policy Research, 2017-2018 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 8-10 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 8-10. 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	40	<0.1	20	<0.1	20	<0.1
Child	6	0.2	5	0.2	1	<0.1
Teen	0	0.0	0	0.0	0	0.0
SRAGE (Self-reported age)						
Adult	328	0.8	154	0.7	174	0.9
Child	0	0.0	0	0.0	0	0.0
Teen	0	0.0	0	0.0	0	0.0

Source: UCLA Center for Health Policy Research, 2017-2018 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 8-9 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 an iterative hot-deck imputation process using stratum and previously hot-decked race and ethnicity variables as the imputation class. Table 8-11 shows the response patterns by interview mode and variable grouping for respondents missing at least one self-reported race or ethnicity value. Table 8-12 shows the response patterns for the self-reported race variables by interview mode.

Table 8-11. Item nonresponse for any self-reported race value and ethnicity by interview mode

Variable and Source of Data	All Modes		Interview Mode[1] Cell		Interview Mode[1] Landline	
	n	pct[2]	n	pct[2]	n	pct[2]
One or more imputed Race values						
Adult	2,616	6.2	1,835	8.3	781	3.9
Child	259	8.1	180	8.8	79	7.0
Teen	83	9.4	49	11.7	34	7.4
SRH (Self-reported Latin ethnicity)						
Adult	269	0.6	149	0.7	120	0.6
Child	24	0.8	12	0.6	12	1.1
Teen	11	1.3	6	1.4	5	1.1

Source: UCLA Center for Health Policy Research, 2017-2018 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 8-9 by mode and person type.

Table 8-12. Item nonresponse for single-response self-reported race by person type and interview mode

Variable and Source of Data	All Modes		Interview Mode[1] Cell		Interview Mode[1] Landline	
	n	pct[2]	n	pct[2]	n	pct[2]
SRW (Self-reported race: White)						
Adult	2,454	5.8	1,747	7.9	707	3.5
Child	250	7.9	174	8.5	76	6.7
Teen	75	8.5	44	10.5	31	6.7
SRAA (Self-reported race: African American)						
Adult	2,454	5.8	1,747	7.9	707	3.5
Child	250	7.9	174	8.5	76	6.7
Teen	75	8.5	44	10.5	31	6.7
SRAI (Self-reported race: American Indian)						
Adult	2,454	5.8	1,747	7.9	707	3.5
Child	250	7.9	174	8.5	76	6.7
Teen	75	8.5	44	10.5	31	6.7

(continued)

Table 8-12. 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		Interview Mode[1] Landline	
	n	pct[2]	n	pct[2]	n	pct[2]
SRAS (Self-reported race: Asian)						
Adult	2,454	5.8	1,747	7.9	707	3.5
Child	250	7.9	174	8.5	76	6.7
Teen	75	8.5	44	10.5	31	6.7
SRPI (Self-reported race: Pacific Islander)						
Adult	2,454	5.8	1,747	7.9	707	3.5
Child	250	7.9	174	8.5	76	6.7
Teen	75	8.5	44	10.5	31	6.7
SRO (Self-reported race: Other)						
Adult	2,454	5.8	1,747	7.9	707	3.5
Child	250	7.9	174	8.5	76	6.7
Teen	75	8.5	44	10.5	31	6.7

Source: UCLA Center for Health Policy Research, 2017-2018 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 8-9 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 Office of Management and Budget (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 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 8-13 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 8-13. 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, 2017-2018 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, child and teen records used reported stratum, SRH, and previously imputed race and ethnicity variables as iterative imputation classes. Records were then classified into the OMB races based on the imputed data. Table 8-14 shows the results of the hot-deck procedure by interview mode, person type and OMBSRREO value.

Table 8-14. 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]
Latino						
Adult	60	0.1	35	0.2	25	0.1
Child	8	0.3	5	0.2	3	0.3
Teen	1	0.1	0	0.0	1	0.2
Non-Latino White Only						
Adult	236	0.6	128	0.6	108	0.5
Child	12	0.4	8	0.4	4	0.4
Teen	11	1.3	7	1.7	4	0.9
Non-Latino African American Only						
Adult	10	<0.1	4	<0.1	6	<0.1
Child	2	<0.1	0	0.0	2	0.2
Teen	2	0.2	1	0.2	1	0.2
Non-Latino American Indian Alaskan Native Only						
Adult	12	<0.1	8	<0.1	4	<0.1
Child	1	<0.1	1	<0.1	0	0.0
Teen	0	0.0	0	0.0	0	0.0
Non-Latino Asian Only						
Adult	24	<0.1	11	<0.1	13	<0.1
Child	2	<0.1	0	0.0	2	0.2
Teen	2	0.2	1	0.2	1	0.2
Non-Latino Pacific Islander Native Hawaiian Only						
Adult	1	<0.1	0	0.0	1	<0.1
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	15	<0.1	11	<0.1	4	<0.1
Child	2	<0.1	1	<0.1	1	<0.1
Teen	2	0.2	2	0.5	0	0.0

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

[1] Unweighted percent of imputed records among respondents in Table 8-9 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 8-15.

Table 8-15. 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, 2017-2018 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, child and teen records used reported region, SRH, and SRAS2 as imputation classes. Table 8-16 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 8-17 shows the results of the hot-deck procedure by interview mode, person type and OMBSRASO value.

Table 8-16. Item nonresponse for single-response self-reported non-Latino Asian ethnicity by person type and interview mode

Single race, Person Type	All Modes		Interview Mode: Cell		Interview Mode: Landline	
	n	pct[1]	n	pct[1]	n	pct[1]
SRCH (OMB Asian ethnicity: Chinese)						
Adult	86	0.2	46	0.2	40	0.2
Child	11	0.4	7	0.3	4	0.4
Teen	2	0.2	1	0.2	1	0.2
SRKR (OMB Asian ethnicity: Korean)						
Adult	86	0.2	46	0.2	40	0.2
Child	11	0.4	7	0.3	4	0.4
Teen	2	0.2	1	0.2	1	0.2
SRPH (OMB Asian ethnicity: Filipino)						
Adult	86	0.2	46	0.2	40	0.2
Child	11	0.4	7	0.3	4	0.4
Teen	2	0.2	1	0.2	1	0.2
SRVT (OMB Asian ethnicity: Vietnamese)						
Adult	86	0.2	46	0.2	40	0.2
Child	11	0.4	7	0.3	4	0.4
Teen	2	0.2	1	0.2	1	0.2
SRASO (OMB Asian ethnicity: Asian Other)						
Adult	86	0.2	46	0.2	40	0.2
Child	11	0.4	7	0.3	4	0.4
Teen	2	0.2	1	0.2	1	0.2
SRJP (OMB Asian ethnicity: Japanese)						
Adult	86	0.2	46	0.2	40	0.2
Child	11	0.4	7	0.3	4	0.4
Teen	2	0.2	1	0.2	1	0.2

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

[1] Unweighted percent of imputed records among respondents in Table 8-9 by mode and person type.

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

OMBSRASO, Person Type	All Modes		Interview Mode: Cell		Interview Mode: Landline	
	n	pct[1]	n	pct[1]	n	pct[1]
Chinese only						
Adult	14	<0.1	8	<0.1	6	<0.1
Child	0	0.0	0	0.0	0	0.0
Teen	0	0.0	0	0.0	0	0.0
Korean only						
Adult	2	<0.1	0	0.0	2	<0.1
Child	1	<0.1	0	0.0	1	<0.1
Teen	0	0.0	0	0.0	0	0.0
Filipino only						
Adult	8	<0.1	8	<0.1	0	0.0
Child	1	<0.1	0	0.0	1	<0.1
Teen	0	0.0	0	0.0	0	0.0
Japanese only						
Adult	2	<0.0	1	<0.0	1	<0.0
Child	0	0.0	0	0.0	0	0.0
Teen	0	0.0	0	0.0	0	0.0
Other Asian ethnicity						
Adult	45	0.1	24	0.1	21	0.1
Child	8	0.3	6	0.3	2	0.2
Teen	1	0.1	0	0.0	1	0.2
Vietnamese only						
Adult	8	0.0	4	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

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

[1] Unweighted percent of imputed records among respondents in Table 8-9 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 8-18). A CART analysis identified the imputation covariates as POVERTY, SRH and OMBSRREO.

Table 8-18. Item nonresponse for self-reported educational attainment of the adult by 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	42,081	99.4	22,076	99.3	20,005	99.5
Imputed values	249	0.6	155	0.7	94	0.5
Total	42,330	100.0	22,231	100.0	20,099	100.0

Source: UCLA Center for Health Policy Research, 2017-2018 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

$$DEFF = \frac{\mathrm{var}_{\pi}(\hat{\theta})}{\mathrm{var}_{stsrs}(\hat{\theta})} \tag{9.1}$$

where $\mathrm{var}_{\pi}(\hat{\theta})$ is the variance estimate for the appropriate CHIS sample design, and $\mathrm{var}_{stsrs}(\hat{\theta})$ is the variance for the stsrs design. Variance for the CHIS sample design, $\mathrm{var}_{\pi}(\hat{\theta})$, 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 with child or teen 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 2017-2018 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.
- **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.
- **Base weight and weight Adjustments**. As discussed in the previous sections of this report, base weights and differential weight adjustments were applied to account for differing selection probabilities across sample frames and sample strata and 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 and to combine ABS with landline/cell samples.

Design effects were computed using SPSS Complex Samples which provides summary statistics and standard errors for complex sample designs. In prior iterations of CHIS, design effects were computed using SUDAAN. 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 2017-2018) or across multiple rounds of a survey using the same sampling design. Thus, differences in DEFF between CHIS 2017-2018 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 24 variables for 2017 and 23 variables for 2018 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 disease, blind/deaf, felt nervous), lifestyle (smoking, number of sexual partners, skipped meals, feel safe), preventive medicine (delayed medical care, usual source of healthcare, number of doctor visits), health insurance (Medicare/Medi-CAL, employer health insurance, other government health plan, prescription coverage), and socioeconomic and demographic variables (income, sexual orientation, marital status, education attainment, U.S. citizenship status). The average DEFT for CHIS 2017 was 1.64 overall and ranging from 0.50 to 1.79 across the reported strata. These values are slightly higher for CHIS 2018 – 1.75 overall and ranging from 0.57 to 2.12 across reported stratum.

A total of 19 variables for 2017 and 18 variables for 2018 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), preventive medicine (usual healthcare location, doctor visits, delayed medical care/medication, access to childcare), and socioeconomic and demographic variables (servings of fruit and vegetables, age). The average DEFT for CHIS 2017 was 1.47 overall and ranged from 0.35 to 1.70 across the reported strata. Design effects were not reported for stratum 44 in 2017 and stratum 36 for 2018 because of the small sample size (n=9 in both cases). These values are slightly higher for CHIS 2018 – 1.55 overall and ranging from 0.31 to 1.78 across reported stratum.

A total of 24 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), lifestyle (smoking, alcohol use, e-cigarette use, had THC, go to the park, park/neighborhood safety concerns, sexually active, walk/bike/skateboard home from school), preventive medicine (usual healthcare location, doctor visits, delayed medical care/medication, physical activity), and socio economic and demographic variables (servings of fruit, vegetables, soda, juice, water). The average DEFT for CHIS 2017 was 1.50 overall and ranged from 0.44 to 1.56 across the reported strata. Note that design effect estimates are only provided for strata with 10 or more teen interviews. These values are slightly higher for CHIS 2018 – 1.55 overall and ranging from 0.44 to 1.50 across reported stratum.

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

	CHIS 2017					CHIS 2018				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
Stratum	Average	Median	Minimum	Maximum	Average	Average	Median	Minimum	Maximum	Average
State	2.70	2.73	1.63	3.48	1.64	3.07	3.19	1.72	3.79	1.75
1 Los Angeles	2.91	3.14	0.12	4.51	1.68	3.19	3.33	1.55	4.56	1.78
2 San Diego	1.71	1.65	0.10	6.85	1.27	1.81	1.87	0.49	4.62	1.33
3 Orange	2.91	3.09	0.17	6.29	1.67	3.72	4.07	0.23	5.75	1.90
4 Santa Clara	3.19	3.47	0.09	8.63	1.73	3.57	3.72	0.34	7.00	1.86
5 San Bernardino	2.89	2.86	0.13	5.17	1.67	2.96	3.18	0.10	4.56	1.69
6 Riverside	2.29	2.45	0.24	5.18	1.47	3.28	3.37	0.92	5.63	1.79
7 Alameda	3.19	3.34	0.59	5.67	1.75	2.88	2.92	0.41	5.71	1.66
8 Sacramento	2.71	2.88	0.06	5.75	1.60	3.29	3.35	0.04	10.27	1.75
9 Contra Costa	2.76	2.92	0.05	8.21	1.61	3.25	3.29	0.31	6.82	1.77
10 Fresno	2.32	2.51	0.05	6.34	1.48	3.52	3.67	0.54	8.34	1.83
11 San Francisco	2.11	2.16	0.18	5.23	1.41	2.19	2.09	0.08	4.15	1.43
12 Ventura	2.50	2.67	0.11	4.59	1.53	2.64	2.76	0.40	5.63	1.59
13 San Mateo	2.94	3.17	0.15	6.91	1.66	3.30	3.46	0.06	6.64	1.75
14 Kern	2.14	2.23	0.13	4.07	1.42	3.54	3.83	0.16	5.35	1.84
15 San Joaquin	3.42	3.76	0.31	7.86	1.79	4.66	4.70	0.78	8.74	2.12
16 Sonoma	2.23	2.34	0.09	5.80	1.44	4.05	3.95	0.70	9.56	1.96
17 Stanislaus	2.45	2.53	0.23	5.59	1.51	2.36	2.38	0.27	6.97	1.49
18 Santa Barbara	1.60	1.55	0.22	3.82	1.25	2.66	2.72	0.22	5.75	1.60
19 Solano	1.89	1.62	0.10	4.52	1.32	1.95	1.94	0.17	3.86	1.36
20 Tulare	2.03	2.12	0.20	4.46	1.38	2.39	2.10	0.06	5.87	1.50
21 Santa Cruz	1.56	1.23	0.33	7.02	1.17	1.28	1.30	0.32	2.60	1.11
22 Marin	1.33	1.55	0.03	4.02	1.10	0.94	1.05	0.10	1.84	0.95
23 San Luis Obispo	0.98	0.94	0.07	3.14	0.95	1.08	1.04	0.07	3.82	1.01
24 Placer	1.83	1.98	0.10	5.05	1.30	2.06	2.07	0.06	3.86	1.38

(continued)

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)

	CHIS 2017					CHIS 2018				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
Stratum	Average	Median	Minimum	Maximum	Average	Average	Median	Minimum	Maximum	Average
State	2.70	2.73	1.63	3.48	1.64	3.07	3.19	1.72	3.79	1.75
25 Merced	1.18	1.22	0.03	3.54	1.05	1.10	0.98	0.02	4.25	0.99
26 Butte	0.90	0.97	0.15	1.93	0.93	1.04	1.05	0.03	2.84	0.98
27 Shasta	0.63	0.67	0.04	1.81	0.76	1.37	1.19	0.28	3.78	1.12
28 Yolo	1.10	1.07	0.05	2.45	1.00	1.11	1.16	0.08	2.12	1.01
29 El Dorado	0.98	0.72	0.13	6.23	0.91	1.01	0.99	0.07	3.58	0.95
30 Imperial	0.55	0.63	0.01	0.99	0.72	0.79	0.78	0.07	2.53	0.86
31 Napa	0.78	0.71	0.03	2.64	0.84	1.32	1.24	0.17	3.48	1.08
32 Kings	0.58	0.61	0.02	1.52	0.74	0.87	0.76	0.05	3.33	0.88
33 Madera	0.60	0.52	0.01	3.60	0.74	0.63	0.70	0.02	2.00	0.75
34 Monterey	1.54	1.55	0.21	4.34	1.22	2.11	2.19	0.12	3.87	1.41
35 Humboldt	0.40	0.39	0.15	0.80	0.62	0.54	0.56	0.05	1.17	0.72
36 Nevada	0.62	0.66	0.06	1.88	0.76	0.43	0.42	0.09	0.77	0.64
37 Mendocino	0.43	0.46	0.04	1.15	0.63	0.45	0.49	0.03	0.87	0.65
38 Sutter	0.48	0.48	0.18	0.87	0.68	0.51	0.48	0.09	1.30	0.69
39 Yuba	0.59	0.68	0.01	1.48	0.73	0.34	0.33	0.03	0.61	0.57
40 Lake	0.54	0.53	0.02	1.75	0.69	0.46	0.34	0.05	1.22	0.64
41 San Benito	0.28	0.22	0.01	1.25	0.50	1.42	0.81	0.09	5.04	1.04
42 Colusa, et al.	0.77	0.75	0.08	2.86	0.84	0.88	0.76	0.12	4.45	0.89
43 Del Norte, et al.	0.94	0.88	0.05	3.17	0.91	0.94	0.86	0.09	2.55	0.94
44 Amador, et al.	1.24	1.16	0.16	3.00	1.05	0.86	0.94	0.01	2.45	0.88

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Note. Imperial County, San Francisco, and AIAN oversamples are included in the design effects computations.

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

	CHIS 2017					CHIS 2018				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
Stratum	Average	Median	Minimum	Maximum	Average	Average	Median	Minimum	Maximum	Average
State	2.18	2.13	1.57	3.12	1.47	2.43	2.45	0.77	3.35	1.55
1 Los Angeles	2.28	2.42	0.70	3.53	1.49	2.38	2.54	0.39	3.66	1.52
2 San Diego	1.32	1.22	0.58	2.15	1.14	1.51	1.68	0.40	3.31	1.21
3 Orange	1.75	1.37	0.43	3.30	1.28	3.09	3.36	0.54	5.40	1.73
4 Santa Clara	2.20	1.63	0.47	4.18	1.45	2.23	2.49	0.90	3.80	1.47
5 San Bernardino	2.57	2.59	1.34	4.09	1.59	1.89	1.94	0.61	2.94	1.36
6 Riverside	1.79	1.78	0.88	3.36	1.32	2.48	2.50	0.50	5.04	1.55
7 Alameda	2.40	2.09	0.13	4.80	1.51	2.17	2.25	0.88	3.19	1.46
8 Sacramento	3.04	2.53	1.47	5.91	1.70	2.52	2.89	0.11	4.48	1.53
9 Contra Costa	1.59	1.79	0.42	3.53	1.22	1.80	1.69	0.77	3.02	1.32
10 Fresno	2.55	2.60	0.25	3.50	1.56	3.39	3.69	0.06	5.41	1.78
11 San Francisco	2.19	1.62	1.04	3.18	1.46	1.60	1.38	0.11	4.23	1.18
12 Ventura	1.77	1.50	0.27	2.92	1.29	2.77	3.11	0.29	4.32	1.60
13 San Mateo	1.05	1.02	0.53	1.53	1.02	1.75	1.88	0.10	3.73	1.24
14 Kern	1.58	1.37	0.29	1.99	1.24	2.65	2.56	1.27	4.67	1.61
15 San Joaquin	2.08	1.59	1.05	2.68	1.43	2.06	2.12	0.26	3.20	1.41
16 Sonoma	0.96	1.02	0.28	1.60	0.96	1.63	1.90	0.26	2.35	1.24
17 Stanislaus	1.48	1.73	0.21	2.60	1.18	2.10	2.22	0.61	3.39	1.41
18 Santa Barbara	1.09	0.95	0.32	2.17	1.01	1.94	2.33	0.38	2.66	1.35
19 Solano	1.26	1.59	0.24	2.45	1.07	1.70	1.72	0.10	3.41	1.21
20 Tulare	2.09	2.33	0.75	3.46	1.41	1.19	0.82	0.14	3.15	1.01
21 Santa Cruz	1.41	1.38	0.25	2.30	1.13	0.79	1.00	0.07	1.40	0.83
22 Marin	1.44	1.43	0.10	2.70	1.08	1.00	0.89	0.27	2.09	0.96
23 San Luis Obispo	1.04	0.91	0.19	2.47	0.97	1.47	1.52	0.24	2.60	1.16
24 Placer	0.66	0.68	0.24	0.91	0.79	1.37	1.12	0.21	3.06	1.12

(continued)

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

	CHIS 2017					CHIS 2018				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
Stratum	Average	Median	Minimum	Maximum	Average	Average	Median	Minimum	Maximum	Average
25 Merced	0.92	0.84	0.14	2.18	0.91	0.98	0.95	0.43	1.70	0.98
26 Butte	0.54	0.56	0.09	0.83	0.72	1.28	1.64	0.04	2.44	1.06
27 Shasta	0.53	0.37	0.13	1.49	0.69	0.43	0.44	0.09	0.88	0.63
28 Yolo	0.57	0.39	0.07	1.37	0.70	0.85	0.90	0.13	1.88	0.89
29 El Dorado	0.59	0.57	0.05	1.21	0.72	0.18	0.17	0.04	0.35	0.40
30 Imperial	0.32	0.35	0.01	0.56	0.54	0.65	0.61	0.25	1.33	0.78
31 Napa	0.62	0.63	0.26	0.92	0.78	0.22	0.21	0.07	0.35	0.46
32 Kings	0.56	0.58	0.05	0.99	0.72	0.38	0.40	0.20	0.59	0.61
33 Madera	0.40	0.55	0.11	0.62	0.63	0.39	0.40	0.06	0.66	0.61
34 Monterey	1.32	1.19	0.54	2.55	1.12	3.06	3.64	0.53	4.58	1.69
35 Humboldt	0.22	0.22	0.07	0.37	0.46	0.42	0.43	0.11	0.81	0.63
36 Nevada	0.27	0.37	0.08	0.51	0.49	**	**	**	**	**
37 Mendocino	0.25	0.25	0.06	0.59	0.47	0.17	0.17	0.06	0.30	0.40
38 Sutter	0.13	0.12	0.03	0.33	0.35	0.10	0.10	0.04	0.21	0.31
39 Yuba	0.16	0.23	0.02	0.25	0.38	0.32	0.26	0.05	0.86	0.53
40 Lake	0.19	0.25	0.02	0.51	0.39	0.15	0.15	0.04	0.28	0.38
41 San Benito	0.27	0.37	0.03	0.65	0.48	0.50	0.34	0.03	1.18	0.64
42 Colusa, et al.	0.32	0.38	0.09	0.44	0.55	0.25	0.23	0.08	0.47	0.49
43 Del Norte, et al.	0.41	0.44	0.10	0.49	0.63	0.50	0.53	0.03	0.92	0.67
44 Amador, et al.	**	**	**	**	**	0.63	0.54	0.07	1.38	0.76

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Note. Design effect estimates are only provided for strata with 10 or more child interviews. Imperial County, San Francisco, and AIAN oversamples are included in the design effects computations.

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

	CHIS 2017					CHIS 2018				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
Stratum	Average	Median	Minimum	Maximum	Average	Average	Median	Minimum	Maximum	Average
State	2.26	2.32	1.60	3.34	1.50	2.42	2.49	0.45	2.88	1.55
1 Los Angeles	2.09	2.32	0.73	3.74	1.43	2.29	2.42	0.39	3.15	1.50
2 San Diego	1.61	1.92	0.06	3.75	1.18	1.80	1.95	0.14	2.89	1.31
3 Orange	2.08	2.54	0.12	4.66	1.35	1.46	1.56	0.10	4.08	1.12
4 Santa Clara	1.80	2.25	0.24	3.20	1.27	1.89	1.89	0.37	3.15	1.35
5 San Bernardino	2.48	2.54	0.58	3.48	1.56	1.99	2.20	0.00	3.60	1.32
6 Riverside	1.95	2.12	0.21	2.76	1.37	1.64	1.86	0.29	3.44	1.24
7 Alameda	1.95	2.22	0.02	2.98	1.32	1.16	0.96	0.13	2.62	1.01
8 Sacramento	0.98	0.69	0.40	2.35	0.96	1.31	1.34	0.03	2.83	1.08
9 Contra Costa	0.91	0.91	0.11	1.82	0.92	**	**	**	**	**
10 Fresno	1.13	0.66	0.10	2.68	0.97	**	**	**	**	**
11 San Francisco	**	**	**	**	**	0.77	0.47	0.05	2.16	0.78
13 San Mateo	0.50	0.51	0.04	0.99	0.67	**	**	**	**	**
14 Kern	2.03	2.19	0.07	3.38	1.36	**	**	**	**	**
23 San Luis Obispo	0.47	0.43	0.09	1.27	0.65	**	**	**	**	**
28 Yolo	**	**	**	**	**	0.25	0.09	0.03	0.59	0.44
29 El Dorado	**	**	**	**	**	0.65	0.85	0.04	0.93	0.76
30 Imperial	0.28	0.14	0.01	0.83	0.44	**	**	**	**	**
35 Humboldt	0.26	0.22	0.02	0.46	0.49	**	**	**	**	**

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.
Note. Design effect estimates are only provided for strata with 10 or more teen interviews. Imperial County, San Francisco, and AIAN oversamples are included in the design effects computations.

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 Chapters 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:

$$\mathrm{v}(\hat{\theta}) = a \sum_{r=1}^{R} \left(\hat{\theta}_{(r)} - \hat{\theta}\right)^2 \qquad (9.2)$$

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 2017-2018 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 2017-2018 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 2015-2016. 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 final base weights shown in Equations (3.1), (3.2) and (3.3):

$$BW_i^{(r)} = \begin{cases} 2 \times BW_i, & \text{if sample unit } i \text{ in variance stratum } s \text{ and variance (group) unit } v \\ 0, & \text{if sample unit } i \text{ in variance stratum } s \text{ and not in variance (group) unit } v \\ BW_i, & \text{if sample unit } i \text{ not in variance stratum } s \end{cases} \quad (9.3)$$

where $s = 1,2, \dots ,40$ 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 two-year replicate weights followed the same process described in Section 9.3.1. The first replicate from 2017 was combined with the first replicate from 2018 using a composite factor specific to that replicate to compute a two-year adjusted base weight. The two-year adjusted base weight for respondent j in replicate i, RBW_{1718ij}, will be calculated as:

$$RBW_{1718ij} = \begin{cases} RBW_{17ij} \times \lambda_{17i} & \text{for 2017 respondents in replicate i} \\ RBW_{18ij} \times (1 - \lambda_{17i}) & \text{for 2018 respondents in replicate i} \end{cases} \qquad (9.4)$$

Where RBW_{17ij} is the final 2017 adjusted replicate base weight for respondent j in replicate i and RBW_{18ij} is the final 2018 adjusted replicate base weight for respondent j in replicate i. λ_{17i} is the proportion of all respondents in replicate i who responded in 2017.

A final adjustment will be made to ensure that each replicate's base weight sums to exactly the target population size of 38,885,450 that was used for the 2018 weighting.

$$FRBW_i = RBW_{1718i} \times \frac{38{,}885{,}450}{\sum_i RBW_{1718i}} \qquad (9.5)$$

Each replicate will then be calibrated to the population control totals that were used for the combined 2017-2018 linear weighting.

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.

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/NSDUH2012MRBAmmended/NSDUHmrbSamplingWgt2012.pdf

Kalton, G. and Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, *19*(2), 8197.

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 DualFrame 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.

National Center for Health Statistics. (2017). Wireless Substitution: State-level Estimates from the National Health Interview Survey, 2016. Division of Health Interview Statistics, National Center for Health Statistics. URL: http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201712.pdf.

National Center for Health Statistics. (2019). Wireless Substitution: State-level Estimates from the National Health Interview Survey, 2017. Division of Health Interview Statistics, National Center for Health Statistics. URL: http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201903.pdf.

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

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.rproject.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 2017-2018 sample design directly related to calculation of the base weights (inverse probability of selection).

Table A-1 contains counts landline and surname sampling frames from CHIS 2017-2018 data collection by design stratum, as well as the associated sample sizes and base weights. Tables A-2, A-3, and A-4 provides the same information (frames, sample sizes, and weights, respectively) as shown in Table A-1 for the cell frames.

Table A-1a. 2017 Landline Frame Sizes, Sample Sizes and Base weights for RDD and Listed Samples

Sample Stratum	Frame size[1]					Sample[2]				Base weight			
	Total	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD
1 Los Angeles	8,801,100	7,577,951	1,091,513	82,903	48,732	73,920	12,983	1,066	970	102.5	84.1	77.8	50.2
2 San Diego	2,753,100	2,503,840	232,345	12,756	4,159	48,053	4,772	294	97	52.1	48.7	43.4	42.9
3 Orange	2,735,500	1,616,270	1,074,094	11,836	33,299	20,483	20,650	254	1,118	78.9	52.0	46.6	29.8
4 Santa Clara	1,540,700	940,505	539,943	24,825	35,427	7,274	5,370	329	605	129.3	100.5	75.5	58.6
5 San Bernardino	1,375,600	1,364,505	0	11,095	0	17,217	0	221	0	79.3		50.2	
6 Riverside	1,480,100	1,469,907	1,176	9,017	0	22,496	21	231	0	65.3	56.0	39.0	
7 Alameda	1,454,400	1,430,304	831	23,265	0	12,050	10	298	0	118.7	83.1	78.1	
8 Sacramento	1,177,900	1,052,277	102,442	20,189	2,991	5,629	609	173	36	186.9	168.2	116.7	83.1
9 Contra Costa	958,300	941,503	0	16,797	0	6,390	0	163	0	147.3		103.0	
10 Fresno	638,400	633,234	0	5,166	0	5,026	0	63	0	126.0		82.0	
11 San Francisco	954,800	924,641	1,317	28,842	0	7,021	11	346	1	131.7	119.7	83.4	83.4
12 Ventura	667,000	663,857	0	3,143	0	9,293	0	87	0	71.4		36.1	
13 San Mateo	796,500	784,766	747	10,987	0	7,357	12	197	0	106.7	62.2	55.8	
14 Kern	511,500	508,770	0	2,730	0	4,472	0	32	0	113.8		85.3	
15 San Joaquin	439,200	434,084	0	5,116	0	3,309	0	59	0	131.2		86.7	
16 Sonoma	458,900	455,988	0	2,912	0	2,036	0	16	0	224.0		182.0	
17 Stanislaus	350,900	349,663	0	1,237	0	2,826	0	19	0	123.7		65.1	
18 Santa Barbara	368,400	366,897	0	1,503	0	2,685	0	14	0	136.6		107.4	
19 Solano	317,800	314,525	0	3,275	0	2,785	0	36	0	112.9		91.0	
20 Tulare	258,300	257,738	0	562	0	3,212	0	7	0	80.2		80.2	
21 Santa Cruz	265,800	263,322	1,014	1,464	0	2,338	9	18	0	112.6	112.6	81.3	
22 Marin	326,600	324,912	0	1,688	0	4,235	0	45	0	76.7		37.5	
23 San Luis Obispo	240,000	239,032	0	968	0	741	0	3	0	322.6		322.6	
24 Placer	320,600	318,609	0	1,991	0	960	0	12	0	331.9		165.9	
25 Merced	129,900	128,961	0	939	0	3,297	0	29	0	39.1		32.4	
26 Butte	169,900	168,946	0	954	0	885	0	5	0	190.9		190.9	

(continued)

Table A-1a. 2017 Landline Frame Sizes, Sample Sizes and Base weights for RDD and Listed Samples (continued)

	Frame size[1]					Sample[2]				Base weight			
Sample Stratum	Total	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD
27 Shasta	144,900	144,467	0	433	0	1,002	0	7	0	144.2		61.8	
28 Yolo	144,300	143,683	0	617	0	699	0	23	0	205.6		26.8	
29 El Dorado	168,800	167,772	0	1,028	0	816	0	10	0	205.6		102.8	
30 Imperial	92,900	92,623	0	277	0	4,675	0	24	0	19.8		11.6	
31 Napa	115,300	114,816	0	484	0	4,746	0	22	0	24.2		22.0	
32 Kings	70,800	70,601	0	199	0	3,895	0	21	0	18.1		9.5	
33 Madera	85,700	85,504	33	163	0	2,623	1	6	0	32.6	32.6	27.2	
34 Monterey	347,900	345,138	0	2,762	0	3,124	0	34	0	110.5		81.2	
35 Humboldt	127,700	127,700	0	0	0	786	0	0	0	162.5			
36 Nevada	108,800	108,012	0	788	0	2,329	0	20	0	46.4		39.4	
37 Mendocino	79,800	79,286	0	514	0	1,697	0	20	0	46.7		25.7	
38 Sutter	59,700	59,351	0	349	0	2,890	0	17	0	20.5		20.5	
39 Yuba	52,500	52,182	0	318	0	2,465	0	42	0	21.2		7.6	
40 Lake	64,100	63,735	0	365	0	1,923	0	14	0	33.1		26.0	
41 San Benito	38,000	37,928	0	72	0	4,732	0	15	0	8.0		4.8	
42 Colusa, et al.	84,300	84,075	0	225	0	749	0	2	0	112.3		112.3	
43 Del Norte, et al.	186,200	185,938	0	262	0	710	0	1	0	261.9		261.9	
44 Amador, et al.	248,200	246,806	0	1,394	0	885	0	5	0	278.9		278.9	

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Note. VKL=Vietnamese/Korean listed; NVKL=not Vietnamese/Korean listed; VKHD=Vietnamese/Korean high density; NVKHD=not Vietnamese/Korean high density.

1 Maximum yield from RDD frame. Frame counts were not available across partitions so for purposes of the base weight computations, the frame was distributed across partitions based on the distribution of the base sample. This lead to one partition (stratum 11, partition 4) where the estimated frame size was zero yet there was one piece of oversample pulled. This case was assigned a base weight equal to its neighboring partition (stratum 11, partition 3).

[2] Includes base sample plus oversamples. Does not include North Imperial county ABS sample.

Table A-1b. 2018 Landline Frame Sizes, Sample Sizes and Base weights for RDD and Listed Samples

Sample Stratum	Frame size[1]					Sample[2]				Base weight			
	Total	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD
1 Los Angeles	8,225,100	6,897,459	1,162,145	88,515	76,981	78,937	37,262	2,779	3,197	87.38	31.19	31.85	24.08
2 San Diego	2,610,400	2,514,421	79,965	13,763	2,252	69,240	5,149	988	292	36.31	15.53	13.93	7.71
3 Orange	2,729,000	1,625,766	1,025,740	17,300	60,195	16,070	31,508	752	3,029	101.17	32.55	23.01	19.87
4 Santa Clara	1,534,300	944,479	525,080	20,637	44,103	5,675	11,707	575	1,414	166.43	44.85	35.89	31.19
5 San Bernardino	1,378,000	1,367,440	73	10,487	0	18,776	1	414	2	72.83	73.00	25.33	71.80
6 Riverside	1,479,600	1,469,001	2,026	8,573	0	18,849	57	327	0	77.94	35.54	26.22	
7 Alameda	1,443,900	1,404,183	1,225	38,390	102	13,753	27	845	4	102.10	45.37	45.43	25.50
8 Sacramento	1,172,900	1,048,123	106,025	15,932	2,820	7,434	1,833	194	49	140.99	57.84	82.12	57.55
9 Contra Costa	956,800	948,346	0	8,454	0	6,843	0	216	0	138.59		39.14	
10 Fresno	635,800	631,814	0	3,986	0	6,816	0	103	0	92.70		38.70	
11 San Francisco	943,400	853,534	1,435	88,040	391	6,544	32	1,406	7	130.43	44.84	62.62	55.86
12 Ventura	663,100	659,605	74	3,421	0	8,869	1	156	0	74.37	74.00	21.93	
13 San Mateo	793,000	776,266	1,296	15,438	0	6,587	32	419	3	117.85	40.50	36.84	112.67
14 Kern	513,800	511,780	0	2,020	0	5,320	0	57	0	96.20		35.44	
15 San Joaquin	440,500	435,986	0	4,514	0	5,022	0	105	0	86.82		42.99	
16 Sonoma	460,200	459,440	0	760	0	3,628	0	12	0	126.64		63.33	
17 Stanislaus	349,300	347,598	0	1,702	0	3,881	0	38	0	89.56		44.79	
18 Santa Barbara	366,900	366,017	0	883	0	3,317	0	17	0	110.35		51.94	
19 Solano	317,000	315,476	0	1,524	0	5,383	0	59	0	58.61		25.83	
20 Tulare	258,400	257,984	0	416	0	4,956	0	35	0	52.05		11.89	
21 Santa Cruz	263,000	260,941	1,336	724	0	4,688	37	23	0	55.66	36.11	31.48	
22 Marin	327,000	325,531	0	1,469	0	5,320	0	48	0	61.19		30.60	
23 San Luis Obispo	239,400	238,511	0	889	0	5,100	0	23	0	46.77		38.65	
24 Placer	325,700	324,521	0	1,179	0	6,329	0	29	0	51.28		40.66	
25 Merced	129,900	129,678	0	222	0	4,674	0	18	0	27.74		12.33	
26 Butte	170,400	169,548	0	852	0	3,979	0	22	0	42.61		38.73	

(continued)

Table A-1b. 2018 Landline Frame Sizes, Sample Sizes and Base weights for RDD and Listed Samples (continued)

	Frame size[1]					Sample[2]				Base weight			
Sample Stratum	Total	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD	NVKL, NVKHD	NVKL, VKHD	VKL, NVKHD	VKL, VKHD
27 Shasta	146,600	145,949	0	651	0	4,034	0	34	0	36.18		19.15	
28 Yolo	143,200	141,635	0	1,565	0	4,977	0	82	0	28.46		19.09	
29 El Dorado	167,600	166,705	0	895	0	5,216	0	36	0	31.96		24.86	
30 Imperial	93,500	93,198	0	302	0	5,547	0	43	0	16.80		7.02	
31 Napa	115,800	115,520	0	280	0	5,773	0	25	0	20.01		11.20	
32 Kings	71,400	71,249	0	151	0	5,646	0	31	0	12.62		4.87	
33 Madera	83,900	83,689	0	211	0	3,170	0	19	0	26.40		11.11	
34 Monterey	344,700	343,271	0	1,429	0	4,804	0	72	0	71.46		19.85	
35 Humboldt	128,200	127,691	0	509	0	3,511	0	15	0	36.37		33.93	
36 Nevada	106,500	106,425	0	75	0	2,841	0	6	0	37.46		12.50	
37 Mendocino	79,200	79,040	0	160	0	2,471	0	11	0	31.99		14.55	
38 Sutter	63,100	62,913	0	187	0	4,377	0	26	0	14.37		7.19	
39 Yuba	49,400	49,335	0	65	0	3,036	0	18	0	16.25		3.61	
40 Lake	64,700	64,584	0	116	0	2,776	0	9	0	23.27		12.89	
41 San Benito	35,800	35,706	0	94	0	3,036	0	22	0	11.76		4.27	
42 Colusa, et al.	82,900	82,606	0	294	0	3,932	0	14	0	21.01		21.00	
43 Del Norte, et al.	159,000	158,617	0	383	0	3,726	0	10	0	42.57		38.30	
44 Amador, et al.	227,900	227,432	0	468	0	4,373	0	14	0	52.01		33.43	

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Note. VKL=Vietnamese/Korean listed; NVKL=not Vietnamese/Korean listed; VKHD=Vietnamese/Korean high density; NVKHD=not Vietnamese/Korean high density.

[1] Maximum yield from RDD frame. Frame counts were not available across partitions so for purposes of the base weight computations, the frame was distributed across partitions based on the distribution of the base sample. This resulted in two partitions (stratum 5, partition 4 and stratum 13, partition 4) where the estimated frame size was zero yet there were pieces of oversample pulled. These cases were assigned the mean stratum base weight.

[2] Includes base sample plus oversamples. Does not include San Francisco and AIAN oversamples.

Table A-2a. 2017 Cell Frame Sizes for RDD and Listed Samples

	Frame Size[1]						
Sample Stratum	Total	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
1 Los Angeles	16,418,719	64,211	103,324	824,961	111,147	5,782,552	9,532,524
2 San Diego	5,250,722	4,971	34,031	37,281	41,870	1,998,081	3,134,488
3 Orange	5,691,073	108,500	82,862	822,671	914,079	1,422,025	2,340,935
4 Santa Clara	3,030,554	81,023	20,864	871,784	90,412	667,313	1,299,157
5 San Bernardino	3,027,623	0	17,795	0	0	944,982	2,064,846
6 Riverside	3,050,620	0	12,904	0	0	951,479	2,086,237
7 Alameda	2,691,321	0	31,298	0	0	937,690	1,722,333
8 Sacramento	2,176,075	0	14,770	0	0	676,197	1,485,108
9 Contra Costa	1,310,049	0	7,472	0	0	564,118	738,459
10 Fresno	1,596,350	0	8,256	0	0	516,458	1,071,636
11 San Francisco	1,911,569	0	21,522	0	0	966,825	923,222
12 Ventura	1,323,697	0	2,963	0	0	402,607	918,128
13 San Mateo	905,516	0	4,236	0	0	297,426	603,854
14 Kern	1,302,301	0	1,303	0	0	415,517	885,481
15 San Joaquin	981,756	0	4,240	0	0	367,262	610,254
16 Sonoma	679,759	0	826	0	0	306,580	372,354
17 Stanislaus	755,934	0	2,394	0	0	280,092	473,449
18 Santa Barbara	601,671	0	2,092	0	0	193,351	406,228
19 Solano	576,453	0	1,105	0	0	268,758	306,590
20 Tulare	591,599	0	431	0	0	183,904	407,263
21 Santa Cruz	349,232	0	961	0	0	102,094	246,178
22 Marin	412,243	0	1,286	0	0	121,840	289,117
23 San Luis Obispo	374,638	0	1,260	0	0	139,324	234,054
24 Placer	517,003	0	2,352	0	0	202,568	312,083
25 Merced	344,867	0	1,010	0	0	114,007	229,850
26 Butte	314,423	0	921	0	0	90,757	222,745
27 Shasta	296,755	0	1,091	0	0	73,534	222,130
28 Yolo	194,217	0	1,022	0	0	55,545	137,650
29 El Dorado	160,005	0	0	0	0	46,201	113,804
30 Imperial	393,690	0	398	0	0	140,774	252,518
31 Napa	141,941	0	130	0	0	38,969	102,842
32 Kings	187,223	0	146	0	0	68,178	118,899
33 Madera	181,856	0	99	0	0	31,201	150,556
34 Monterey	597,540	0	1,624	0	0	194,091	401,825
35 Humboldt	191,044	0	417	0	0	38,682	151,945

(continued)

Table A-2a. 2017 Cell Frame Sizes for RDD and Listed Samples (continued)

Sample Stratum	Frame Size[1]						
	Total	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
36 Nevada	115,658	0	0	0	0	23,685	91,973
37 Mendocino	125,802	0	0	0	0	21,198	104,604
38 Sutter	234,918	0	55	0	0	93,779	141,083
39 Yuba	16,199	0	0	0	0	12,959	3,240
40 Lake	65,433	0	19	0	0	13,056	52,358
41 San Benito	84,002	0	147	0	0	19,901	63,954
42 Colusa, et al.	90,024	0	0	0	0	22,486	67,538
43 Del Norte, et al.	167,377	0	0	0	0	27,264	140,113
44 Amador, et al.	217,228	0	389	0	0	70,337	146,503

Source: UCLA Center for Health Policy Research, 2017 California Health Interview Survey.

Note. VKL=Vietnamese/Korean listed; NVKL=not Vietnamese/Korean listed; VKHD=Vietnamese/Korean high density; NVKHD=not Vietnamese/Korean high density; L=listed; NL=not listed.

[1] Maximum yield from RDD frame, plus maximum out-of-area yield. Frame counts were not available across partitions so for purposes of the base weight computations, the frame was distributed across partitions based on the distribution of the base sample. This lead to one partition (stratum 11, partition 4) where the estimated frame size was zero yet there was one piece of oversample pulled. This case was assigned a base weight equal to its neighboring partition (stratum 11, partition 3).

Table A-2b. 2018 Cell Frame Sizes for RDD and Listed Samples

Sample Stratum	Frame Size[1]						
	Total	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
1 Los Angeles	16,815,810	25,014	150,540	353,963	1,272,508	3,744,537	11,269,248
2 San Diego	5,292,397	2,193	19,686	20,293	100,297	1,280,392	3,869,538
3 Orange	5,746,607	73,006	63,215	564,891	2,085,883	945,033	2,014,580
4 Santa Clara	3,050,691	50,908	24,505	523,922	864,230	419,000	1,168,126
5 San Bernardino	3,064,520	0	13,101	0	0	808,326	2,243,093
6 Riverside	2,999,010	0	7,156	0	0	649,445	2,342,409
7 Alameda	2,742,270	0	48,402	0	0	743,940	1,949,927
8 Sacramento	2,206,933	0	23,973	0	0	552,805	1,630,155
9 Contra Costa	1,296,653	0	4,351	0	0	371,816	920,486
10 Fresno	1,591,682	0	2,473	0	0	292,489	1,296,720
11 San Francisco	1,925,204	0	111,595	0	0	590,940	1,222,669
12 Ventura	1,350,191	0	1,608	0	0	369,156	979,427
13 San Mateo	915,947	0	10,205	0	0	221,843	683,899
14 Kern	1,323,650	0	939	0	0	234,700	1,088,011
15 San Joaquin	993,051	0	3,989	0	0	252,727	736,335
16 Sonoma	658,022	0	967	0	0	123,984	533,071
17 Stanislaus	760,173	0	694	0	0	215,706	543,773
18 Santa Barbara	597,898	0	1,466	0	0	157,264	439,168
19 Solano	581,768	0	1,035	0	0	146,793	433,940
20 Tulare	591,720	0	884	0	0	113,037	477,798
21 Santa Cruz	342,386	0	191	0	0	48,705	293,490
22 Marin	417,437	0	1,413	0	0	92,694	323,330
23 San Luis Obispo	384,803	0	389	0	0	80,829	303,585
24 Placer	527,330	0	772	0	0	119,770	406,788
25 Merced	347,975	0	480	0	0	61,863	285,632
26 Butte	316,537	0	358	0	0	65,802	250,377
27 Shasta	306,931	0	270	0	0	60,985	245,675
28 Yolo	197,344	0	392	0	0	44,785	152,166
29 El Dorado	162,155	0	65	0	0	37,737	124,353
30 Imperial	402,776	0	364	0	0	54,394	348,018
31 Napa	144,008	0	127	0	0	36,560	107,320
32 Kings	188,274	0	97	0	0	37,352	150,825
33 Madera	182,898	0	79	0	0	23,777	159,042
34 Monterey	631,739	0	1,437	0	0	103,006	527,296
35 Humboldt	193,140	0	64	0	0	23,579	169,497

(continued)

Table A-2b. 2018 Cell Frame Sizes for RDD and Listed Samples (continued)

Sample Stratum	Frame Size[1]						
	Total	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
36 Nevada	115,728	0	56	0	0	21,011	94,661
37 Mendocino	126,848	0	0	0	0	15,577	111,271
38 Sutter	234,972	0	346	0	0	52,920	181,706
39 Yuba	16,833	0	35	64	0	16,719	15
40 Lake	65,661	0	33	0	0	9,919	55,709
41 San Benito	86,453	0	45	0	0	15,162	71,246
42 Colusa, et al.	88,046	0	35	0	0	25,398	62,613
43 Del Norte, et al.	165,180	0	36	0	0	21,732	143,412
44 Amador, et al.	221,532	0	126	0	0	32,079	189,327

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Note. VKL=Vietnamese/Korean listed; NVKL=not Vietnamese/Korean listed; VKHD=Vietnamese/Korean high density; NVKHD=not Vietnamese/Korean high density; L=listed; NL=not listed.

[1] Maximum yield from RDD frame, plus maximum out-of-area yield. Frame counts were not available across partitions so for purposes of the base weight computations, the frame was distributed across partitions based on the distribution of the base sample.

Table A-3a. 2017 Cell Sample Sizes for RDD and Listed Samples

Sample Stratum	Sample Size[2]					
	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
1 Los Angeles	261	583	3,225	341	20,255	29,246
2 San Diego	30	265	227	234	12,595	16,395
3 Orange	410	501	3,029	2,939	5,231	6,300
4 Santa Clara	324	260	3,523	264	2,787	3,736
5 San Bernardino	0	85	0	0	3,210	6,730
6 Riverside	0	113	0	0	5,022	10,347
7 Alameda	0	162	0	0	3,105	5,503
8 Sacramento	0	129	0	0	3,173	6,938
9 Contra Costa	0	56	0	0	4,215	4,744
10 Fresno	0	50	0	0	1,826	3,764
11 San Francisco	0	105	0	0	4,256	3,303
12 Ventura	0	25	0	0	1,347	2,789
13 San Mateo	0	41	0	0	1,993	3,421
14 Kern	0	6	0	0	1,669	3,399
15 San Joaquin	0	28	0	0	1,132	1,871
16 Sonoma	0	3	0	0	1,114	1,353
17 Stanislaus	0	13	0	0	1,521	2,571
18 Santa Barbara	0	15	0	0	1,116	2,330
19 Solano	0	16	0	0	2,663	2,496
20 Tulare	0	9	0	0	855	1,889
21 Santa Cruz	0	7	0	0	745	1,794
22 Marin	0	12	0	0	1,140	2,698
23 San Luis Obispo	0	5	0	0	555	929
24 Placer	0	9	0	0	603	929
25 Merced	0	18	0	0	1,537	2,504
26 Butte	0	8	0	0	394	967
27 Shasta	0	7	0	0	337	1,018
28 Yolo	0	8	0	0	435	1,078
29 El Dorado	0	0	0	0	449	1,106
30 Imperial	0	8	0	0	1,995	2,540
31 Napa	0	4	0	0	1,322	3,159
32 Kings	0	5	0	0	1,703	2,445
33 Madera	0	3	0	0	897	3,040
34 Monterey	0	10	0	0	1,306	2,474
35 Humboldt	0	3	0	0	278	1,092

(continued)

Table A-3a. 2017 Cell Sample Sizes for RDD and Listed Samples (continued)

Sample Stratum	Sample Size[2]					
	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
36 Nevada	0	0	0	0	651	2,528
37 Mendocino	0	0	0	0	627	3,089
38 Sutter	0	3	0	0	2,262	2,553
39 Yuba	0	0	0	0	2,543	18
40 Lake	0	1	0	0	721	2,707
41 San Benito	0	10	0	0	1,340	3,474
42 Colusa, et al.	0	0	0	0	278	835
43 Del Norte, et al.	0	0	0	0	187	961
44 Amador, et al.	0	3	0	0	546	1,131

Source: UCLA Center for Health Policy Research, 2017 California Health Interview Survey.

Note. VKL=Vietnamese/Korean listed; NVKL=not Vietnamese/Korean listed; VKHD=Vietnamese/Korean high density; NVKHD=not Vietnamese/Korean high density; L=listed; NL=not listed.

[2] Includes base sample plus oversamples. Does not include North Imperial county ABS sample.

Table A-3b. 2018 Cell Sample Sizes for RDD and Listed Samples

	Sample Size[2]					
Sample Stratum	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
1 Los Angeles	332	1,905	4,634	21,108	44,133	98,667
2 San Diego	59	765	543	2,663	31,448	82,949
3 Orange	630	1,253	4,630	19,280	7,566	9,466
4 Santa Clara	619	640	6,844	16,003	4,076	6,769
5 San Bernardino	0	215	0	0	8,136	20,546
6 Riverside	0	136	0	0	7,931	24,876
7 Alameda	0	665	0	0	8,246	20,143
8 Sacramento	0	219	0	0	3,890	10,268
9 Contra Costa	0	129	0	0	3,674	7,828
10 Fresno	0	24	0	0	1,653	6,291
11 San Francisco	0	1,048	0	0	5,199	9,532
12 Ventura	0	29	0	0	3,390	7,917
13 San Mateo	0	175	0	0	3,614	8,712
14 Kern	0	11	0	0	1,828	6,949
15 San Joaquin	0	69	0	0	2,989	8,123
16 Sonoma	0	8	0	0	760	2,756
17 Stanislaus	0	31	0	0	4,497	10,976
18 Santa Barbara	0	34	0	0	2,564	6,892
19 Solano	0	40	0	0	4,137	11,319
20 Tulare	0	20	0	0	1,385	5,402
21 Santa Cruz	0	10	0	0	1,347	7,689
22 Marin	0	123	0	0	5,377	18,529
23 San Luis Obispo	0	19	0	0	2,008	7,016
24 Placer	0	27	0	0	2,274	7,377
25 Merced	0	37	0	0	2,535	10,721
26 Butte	0	15	0	0	1,345	4,897
27 Shasta	0	25	0	0	2,759	10,905
28 Yolo	0	51	0	0	3,576	12,021
29 El Dorado	0	12	0	0	3,538	11,461
30 Imperial	0	17	0	0	2,073	12,425
31 Napa	0	22	0	0	3,394	9,273
32 Kings	0	18	0	0	3,596	13,955
33 Madera	0	15	0	0	2,263	12,067
34 Monterey	0	34	0	0	1,189	5,503
35 Humboldt	0	2	0	0	769	5,262

(continued)

Table A-3b. 2018 Cell Sample Sizes for RDD and Listed Samples (continued)

Sample Stratum	Sample Size[2]					
	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
36 Nevada	0	10	0	0	2,270	10,146
37 Mendocino	0	0	0	0	929	6,479
38 Sutter	0	88	0	0	9,910	32,540
39 Yuba	0	8	17	0	3,703	3
40 Lake	0	3	0	0	976	5,083
41 San Benito	0	44	0	0	9,877	45,514
42 Colusa, et al.	0	2	0	0	1,479	3,587
43 Del Norte, et al.	0	1	0	0	642	3,966
44 Amador, et al.	0	5	0	0	816	4,515

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Note. VKL=Vietnamese/Korean listed; NVKL=not Vietnamese/Korean listed; VKHD=Vietnamese/Korean high density; NVKHD=not Vietnamese/Korean high density; L=listed; NL=not listed.

[2] Includes base sample plus oversamples. Does not include San Francisco and AIAN oversamples.

Table A-4a. 2017 Cell Frame Base weights for RDD and Listed Samples

Sample Stratum	Base weight					
	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
1 Los Angeles	246.0	177.2	255.8	325.9	285.5	325.9
2 San Diego	165.7	128.4	164.2	178.9	158.6	191.2
3 Orange	264.6	165.4	271.6	311.0	271.8	371.6
4 Santa Clara	250.1	80.2	247.5	342.5	239.4	347.7
5 San Bernardino		209.4			294.4	306.8
6 Riverside		114.2			189.5	201.6
7 Alameda		193.2			302.0	313.0
8 Sacramento		114.5			213.1	214.1
9 Contra Costa		133.4			133.8	155.7
10 Fresno		165.1			282.8	284.7
11 San Francisco		205.0			227.2	279.5
12 Ventura		118.5			298.9	329.2
13 San Mateo		103.3			149.2	176.5
14 Kern		217.1			249.0	260.5
15 San Joaquin		151.4			324.4	326.2
16 Sonoma		275.2			275.2	275.2
17 Stanislaus		184.1			184.1	184.1
18 Santa Barbara		139.5			173.3	174.3
19 Solano		69.1			100.9	122.8
20 Tulare		47.9			215.1	215.6
21 Santa Cruz		137.2			137.0	137.2
22 Marin		107.2			106.9	107.2
23 San Luis Obispo		251.9			251.0	251.9
24 Placer		261.3			335.9	335.9
25 Merced		56.1			74.2	91.8
26 Butte		115.2			230.3	230.3
27 Shasta		155.9			218.2	218.2
28 Yolo		127.7			127.7	127.7
29 El Dorado					102.9	102.9
30 Imperial		49.7			70.6	99.4
31 Napa		32.6			29.5	32.6
32 Kings		29.2			40.0	48.6
33 Madera		33.0			34.8	49.5
34 Monterey		162.4			148.6	162.4
35 Humboldt		139.1			139.1	139.1

(continued)

Table A-4a. 2017 Cell Frame Base weights for RDD and Listed Samples (continued)

Sample Stratum	Base weight					
	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
36 Nevada					36.4	36.4
37 Mendocino					33.8	33.9
38 Sutter		18.4			41.5	55.3
39 Yuba					6.3	6.3
40 Lake		19.3			18.1	19.3
41 San Benito		14.7			14.9	18.4
42 Tehama-Glenn-Colusa					80.9	80.9
43 Del Norte-Siskiyou-Lassen-Trinity-Modoc-Plumas-Sierra					145.8	145.8
44 Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine		129.5			128.8	129.5

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Note. VKL=Vietnamese/Korean listed; NVKL=not Vietnamese/Korean listed; VKHD=Vietnamese/Korean high density; NVKHD=not Vietnamese/Korean high density; L=listed; NL=not listed.

Table A-4b. 2018 Cell Frame Base weights for RDD and Listed Samples

	Base weight					
Sample Stratum	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
1 Los Angeles	75.3	79.0	76.4	60.3	84.8	114.2
2 San Diego	37.2	25.7	37.4	37.7	40.7	46.6
3 Orange	115.9	50.5	122.0	108.2	124.9	212.8
4 Santa Clara	82.2	38.3	76.6	54.0	102.8	172.6
5 San Bernardino		60.9			99.4	109.2
6 Riverside		52.6			81.9	94.2
7 Alameda		72.8			90.2	96.8
8 Sacramento		109.5			142.1	158.8
9 Contra Costa		33.7			101.2	117.6
10 Fresno		103.0			176.9	206.1
11 San Francisco		106.5			113.7	128.3
12 Ventura		55.4			108.9	123.7
13 San Mateo		58.3			61.4	78.5
14 Kern		85.4			128.4	156.6
15 San Joaquin		57.8			84.6	90.6
16 Sonoma		120.9			163.1	193.4
17 Stanislaus		22.4			48.0	49.5
18 Santa Barbara		43.1			61.3	63.7
19 Solano		25.9			35.5	38.3
20 Tulare		44.2			81.6	88.4
21 Santa Cruz		19.1			36.2	38.2
22 Marin		11.5			17.2	17.4
23 San Luis Obispo		20.5			40.3	43.3
24 Placer		28.6			52.7	55.1
25 Merced		13.0			24.4	26.6
26 Butte		23.9			48.9	51.1
27 Shasta		10.8			22.1	22.5
28 Yolo		7.7			12.5	12.7
29 El Dorado		5.4			10.7	10.9
30 Imperial		21.4			26.2	28.0
31 Napa		5.8			10.8	11.6
32 Kings		5.4			10.4	10.8
33 Madera		5.3			10.5	13.2
34 Monterey		42.3			86.6	95.8
35 Humboldt		32.0			30.7	32.2

(continued)

Table A-4b. 2018 Cell Frame Base weights for RDD and Listed Samples (continued)

	Base weight					
Sample Stratum	VKL, VKHD, L	VKL, NVKHD, L	NVKL, VKHD, L	NVKL, VKHD, NL	NVKL, NVKHD, L	NVKL, NVKHD, NL
36 Nevada		5.6			9.3	9.3
37 Mendocino					16.8	17.2
38 Sutter		3.9			5.3	5.6
39 Yuba		4.4	3.8		4.5	5.0
40 Lake		11.0			10.2	11.0
41 San Benito		1.0			1.5	1.6
42 Colusa, et al.		17.5			17.2	17.5
43 Del Norte, et al.		36.0			33.9	36.2
44 Amador, et al.		25.2			39.3	41.9

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Note. VKL=Vietnamese/Korean listed; NVKL=not Vietnamese/Korean listed; VKHD=Vietnamese/Korean high density; NVKHD=not Vietnamese/Korean high density; L=listed; NL=not listed.

APPENDIX B – Summary Statistics for Weights and Weight Adjustments

Appendix B includes summary statistics on the CHIS 2017-2018 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	CHIS 2017 Sampling Frame			CHIS 2018 Sampling Frame	
	Landline	Cell	ABS	Landline	Cell
1. Base weight					
1.1 Sample size	368,311	283,526	5,252	562,406	1,088,227
1.2 Sum of weights	31,711,183	59,646,679	10,746	31,027,830	60,634,570
1.3 Coefficient of variation	49.3	46.6	0.0	63.6	86.5
2. Unknown residential status adjustment					
2.1 Sample size					
a. Known residential status	100,418	173,634	2,911	109,273	347,456
b. Unknown residential status	267,893	109,892	2,341	453,133	740,771
2.2 Sum of weights	10,271,725	41,317,408	10,746	9,225,819	31,104,261
2.3 Coefficient of variation	53.5	46.6	0.0	68.0	86.5
2.4 Mean non-zero adjustment	0.90	1.03	1.48	1.60	1.60
3. Screener nonresponse adjustment					
3.1 Sample size					
a. Screener respondents	16,054	16,916	744	11,488	13,855
b. Screener nonrespondents	84,364	156,718	2,167	97,785	333,601
3.2 Sum of weights	10,271,725	41,317,408	10,746	6,986,984	12,923,401
3.3 Coefficient of variation	81.9	58.2	23.6	93.9	92.6
3.4 Mean non-zero adjustment	5.88	10.37	7.06	6.47	8.96
3. Asian non-consent adjustment					
3.1 Sample size					
a. Asian cases with reconsent				234	122
b. Asian cases no reconsent				323	196
c. Non-Asian cases				10,931	13,537
3.2 Sum of weights				6,999,184	12,911,201
3.3 Coefficient of variation				93.5	94.0
3.4 Mean non-zero adjustment				1.03	1.01
4. Listed 65+ adjustment					
4.1 Sample size	16,054	16,916	744	11,165	13,659
4.2 Sum of weights	13,652,074	47,821,289	10,746	8,350,558	13,639,370
4.3 Coefficient of variation	115.2	92.4	23.6	116.8	105.0
4.4 Mean non-zero adjustment	1.36	1.15	1.00	1.21	1.06

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Table B-2. Extended adult interview weighting adjustments by sample type

Survey Weight Statistics	CHIS 2017 Sampling Frame			CHIS 2018 Sampling Frame	
	Landline	Cell	ABS	Landline	Cell
1. Number of adults adjustment					
1.1 Sample size	16,054	16,916	744	11,165	13,659
1.2 Sum of weights	24,748,382	47,821,289	20,997	14,570,943	13,639,370
1.3 Coefficient of variation	117.4	92.4	45.2	121.6	105.0
1.4 Mean non-zero adjustment factor	1.89	1.00	1.94	1.78	1.00
2. Adult nonresponse adjustment					
2.1 Sample size					
a. Adult respondents	10,020	10,794	339	9,744	11,433
b. Adult nonrespondents	6,034	6,122	405	1,421	2,226
2.2 Sum of weights	24,756,575	48,015,592	19,607	14,589,828	13,620,486
2.3 Coefficient of variation	134.0	95.0	73.9	125.1	107.6
2.4 Mean non-zero adjustment factor	1.73	1.64	2.07	1.16	1.18
3. Phone use calibration					
3.1 Sample size	10,020	10,794	339	9,744	11,433
3.2 Sum of weights	13,962,000	27,246,520	19,607	14,077,145	27,471,223
3.3 Coefficient of variation	140.9	106.3	73.9	131.9	119.7
3.4 Mean non-zero adjustment factor	0.56	0.56	1.00	0.95	1.99
4. Composite of landline and cell samples					
4.1 Sample size	10,020	10,794	339	9,744	11,433
4.2 Sum of weights	10,149,236	18,535,406	19,607	10,332,449	18,585,535
4.3 Coefficient of variation	135.8	97.3	73.9	129.2	111.0
4.4 Mean non-zero adjustment factor	0.77	0.78	1.00	0.76	0.80
5. Composite of North Imperial county ABS sample					
5.1 Sample size	10,020	10,794	339		
5.2 Sum of weights	10,128,218	18,479,869	11,536		
5.3 Coefficient of variation	136.2	97.7	77.8		
5.4 Mean non-zero adjustment factor	0.99	0.99	0.59		
6. Pre-calibration trimming					
6.1 Number of records trimmed	574	289	0	0	0
6.2 Sum of weights	9,852,697	17,856,579	11,536	10,332,449	18,585,535
6.3 Coefficient of variation	122.9	81.8	77.8	129.2	111.0
7. Final calibration adjustment					
7.1 Sample size	10,020	10,794	339	9,744	11,433
7.2 Sum of weights	9,416,140	20,021,008	18,549	9,630,557	20,068,062
7.3 Coefficient of variation	165.3	110.4	140.7	172.2	134.2
7.4 Mean weight	939.7	1,854.8	54.7	988.4	1,755.3

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Table B-3. Extended child interview weighting adjustments by sample type

Survey Weight Statistics	CHIS 2017 Sampling Frame Landline	CHIS 2017 Sampling Frame Cell	CHIS 2017 Sampling Frame ABS	CHIS 2018 Sampling Frame Landline	CHIS 2018 Sampling Frame Cell
1. Adult nonresponse adjustment					
1.1 Sample size					
a. Adult respondents	769	1,730	74	9,968	11,438
b. Adult nonrespondents	137	57	11	1,520	2,417
1.2 Sum of weights	13,747,310	47,730,844	5,955	8,615,746	13,374,182
1.3 Coefficient of variation	107.0	95.6	22.1	117.8	106.7
1.4 Mean non-zero adjustment factor	1.65	1.64	1.14	1.14	1.17
2. Base weight					
2.1 Sample size	769	1,730	74	756	1,847
2.2 Sum of weights	822,947	7,700,890	1,686	745,663	3,363,303
2.3 Coefficient of variation	128.1	109.5	71.2	117.1	138.2
3. Child nonresponse adjustment					
3.1 Sample size					
a. Child respondents	512	1,046	42	573	1,013
b. Child nonrespondents	257	684	32	183	834
3.2 Sum of weights	1,007,507	7,516,133	1,883	1,030,370	3,078,596
3.3 Coefficient of variation	137.3	101.4	72.8	121.4	149.2
3.4 Mean non-zero adjustment factor	1.82	1.75	2.16	1.95	1.80
4. Phone use calibration					
4.1 Sample size	512	1,046	42	573	1,013
4.2 Sum of weights	2,343,828	5,795,647	1,883	2,327,231	5,754,608
4.3 Coefficient of variation	136.8	110.5	72.8	121.3	157.4
4.4 Mean non-zero adjustment factor	2.33	0.77	1.00	2.26	1.90
5. Composite of landline and cell					
5.1 Sample size	512	1,046	42	573	1,013
5.2 Sum of weights	1,715,005	4,240,900	1,857	1,771,550	4,187,324
5.3 Coefficient of variation	135.1	105.4	73.8	122.0	152.5
5.4 Mean non-zero adjustment factor	0.74	0.84	0.98	0.76	0.85
6. Composite of North Imperial county ABS sample					
6.1 Sample size	512	1,046	42		
6.2 Sum of weights	1,707,964	4,227,907	1,093		
6.3 Coefficient of variation	135.9	105.8	73.5		
6.4 Mean non-zero adjustment factor	0.99	0.99	0.59		
7. Pre-calibration trimming					
7.1 Number of records trimmed	3	1	0	0	0
7.2 Sum of weights	1,621,259	4,226,849	1,093	1,771,550	4,187,324
7.3 Coefficient of variation	113.4	105.8	73.5	122.0	152.5
8. Final calibration adjustment					
8.1 Sample size	512	1,046	42	573	1,013
8.2 Sum of weights	1,643,331	4,442,915	1,619	1,930,386	4,114,371
8.3 Coefficient of variation	115.3	98.9	70.5	115.8	121.4
8.4 Mean weight	3,209.6	4,247.5	38.5	3,368.9	4,061.6

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.

Table B-4. Extended teen interview weighting adjustments by sample type

Survey Weight Statistics	CHIS 2017 Sampling Frame Landline	CHIS 2017 Sampling Frame Cell	CHIS 2017 Sampling Frame ABS	CHIS 2018 Sampling Frame Landline	CHIS 2018 Sampling Frame Cell
1. Base weight					
1.1 Sample size	802	1,231	46	774	1,273
1.2 Sum of weights	889,723	5,870,583	965	926,839	2,349,481
1.3 Coefficient of variation	96.8	113.5	52.5	110.7	118.4
2. Teen nonresponse adjustment					
2.1 Sample size					
a. Teen respondents	211	222	15	233	199
b. Teen nonrespondents	591	1,009	31	541	1,074
2.2 Sum of weights	1,432,474	5,062,185	1,234	1,553,599	1,722,722
2.3 Coefficient of variation	107.0	80.3	39.8	97.4	111.5
2.4 Mean non-zero adjustment factor	5.86	5.8	3.21	6.13	5.75
3. Phone use calibration					
3.1 Sample size	211	222	15	233	199
3.2 Sum of weights	1,207,814	2,986,594	1,234	1,209,698	2,991,254
3.3 Coefficient of variation	107.0	80.6	39.8	96.8	110.3
3.4 Mean non-zero adjustment factor	0.84	0.59	1.00	0.78	1.74
4. Composite of landline and cell samples					
4.1 Sample size	211	222	15	233	199
4.2 Sum of weights	871,277	2,195,779	1,234	710,002	2,412,685
4.3 Coefficient of variation	107.2	103.5	39.8	103.1	114.4
4.4 Mean non-zero adjustment factor	0.73	0.72	1.00	0.59	0.78
5. Composite of North Imperial county ABS sample					
5.1 Sample size	211	222	15		
5.2 Sum of weights	869,683	2,182,133	713		
5.3 Coefficient of variation	107.5	104.5	39.8		
5.4 Mean non-zero adjustment factor	0.98	0.99	0.58		
6. Final calibration adjustment					
6.1 Sample size	211	222	15	233	199
6.2 Sum of weights	803,076	2,333,297	805	928,251	2,213,823
6.3 Coefficient of variation	119.1	87.7	55.5	115.5	100.0
6.4 Mean weight	3,806.05	10,510.35	53.70	3,983.91	11,124.74

Source: UCLA Center for Health Policy Research, 2017-2018 California Health Interview Survey.