california
health
interview
survey

CHIS 2019-2020 Methodology Report Series

Report 1

Sample Design

September 2021

UCLA CENTER FOR
HEALTH POLICY RESEARCH

CALIFORNIA HEALTH INTERVIEW SURVEY

CHIS 2019-2020 METHODOLOGY SERIES

REPORT 1

SAMPLE DESIGN

SEPTEMBER 2021

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

www.chis.ucla.edu

This report provides analysts with information about the sampling methods used for CHIS 2019-2020, including both the household and person (within household) sampling. This report also provides a discussion on achieved sample size and how it compares to the planned sample size.

Suggested citation:

California Health Interview Survey. *CHIS 2019-2020 Methodology Series: Report 1 - Sample Design*. Los Angeles, CA: UCLA Center for Health Policy Research, 2021.

The California Health Interview Survey is a collaborative project of the UCLA Center for Health Policy Research with multiple funding sources. Funding for CHIS 2019-2020 came from the following 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 Mental Health Services Authority, the California Tobacco Control Program, the California Wellness Foundation, First 5 California, Kaiser Permanente, the Long Term Services and Supports Content Development Project, and San Diego County Health and Human Services Agency.

PREFACE

Sample Design is the first in a series of methodological reports describing the 2019-2020 California Health Interview Survey (CHIS 2019-2020). The other reports are listed below.

CHIS is a collaborative project of the University of California, Los Angeles (UCLA) Center for Health Policy Research with multiple funding sources from public, private, and non-profit organizations. SSRS was responsible for data collection and the preparation of five methodological reports from the 2019-2020 survey. The survey examines public health and health care access issues in California. The survey is the largest state health survey ever undertaken in the United States.

Methodological Report Series for CHIS 2019-2020

The methodological reports for CHIS 2019-2020 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 1: Sample Design (this report) describes the procedures used to design and select the sample for CHIS 2019-2020. An appropriate sample design is a feature of a successful survey, and CHIS 2019-2020 presented many issues that had to be addressed at the design stage. This report explains why the design features of CHIS were selected and presents the alternatives that were considered and provides analysts information about the sampling methods used for both the household and person (within household) sampling. In general terms, once a household was sampled, an adult within that household was sampled. If there were children and/or adolescents in the household, one child and/or one adolescent was eligible for sampling. This report also provides a discussion on achieved sample size and how it compares to the planned sample size.

The purposes of this report are:

- To serve as a reference for researchers using CHIS 2019-2020 data;
- To document data collection procedures so that future iterations of CHIS, or other similar surveys, can replicate those procedures if desired;
- To describe lessons learned from the data collection experience and make recommendations for improving future surveys; and
- To evaluate the level of effort required for the various kinds of data collection undertaken.

For further methodological details not covered in this report, refer to the other methodological reports in the series at http://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 2019-2020 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 2019-2020.

- 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 multimode (web and 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 multiple funding sources from public, private, and non-profit organizations. 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 in Sampling and Data Collection Methodology

Starting in 2019-2020, the CHIS transitioned from a dual-frame landline/cellphone random digit dial (RDD) methodology to an address-based sample (ABS) methodology with multimode data collection that takes place on the web or by telephone. The CHIS research team deemed this change necessary due to decreasing response to telephone surveys, the improved geographical precision available for stratification when using the US Postal Service Delivery Sequence file of addresses as a sampling frame, and the lower cost of a study where the majority of interviews are completed online.

Prior to launching data collection in 2019, CHIS conducted two experiments in 2018 to test the effectiveness of an ABS mail push-to-web design with a telephone nonresponse follow-up. The first experiment was limited to three counties (Los Angeles, Santa Clara, and Tulare) to achieve a preliminary assessment of the efficacy of the proposed design (see Wells et al., 2018). Following the initial positive results from that test, a statewide pilot test was conducted in the late 2018 implementing a number of additional experiments and improvements based on the previous lessons learned (see Wells et al., 2019). Given that these additional improvements resulted in higher response and reductions in cost compared to maintaining the 2017-2018 design, CHIS committed to transitioning to the new design for the 2019-2020 cycle.

For CHIS 2019-2020, respondents are invited to either complete the survey online or call in to be interviewed by a member of the SSRS interviewing staff. Respondents receive an initial invitation letter with a $2.00 pre-incentive. This is followed by a reminder postcard and, in 2019, a final certified mail letter for all nonresponders[1]. In 2020, the certified mail letter was replaced with a standard letter and final postcard. Where addresses can be matched to a listed telephone number, the nonresponding households are also called six times to attempt to complete an interview before the sampled household is considered to be a resolved nonresponse.

See more about what's new in the 2019-2020 CHIS sampling and data collection here:

[1] For the last 2019 mailing, the certified letter was replaced with a standard letter.

https://chis.ucla.edu/chis/design/Documents/whats-new-chis-2019-2020.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 2019-2020 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 and year.

1.3 Sample Design Objectives

The CHIS 2019-2020 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 an address-based sample design. For the ABS 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 cycle—to provide health estimates for adults at the local level.

In addition, for CHIS 2019-2020, statistical modeling was used to determine the likelihood that specific targeted groups of interest for oversampling resided at addresses in the sample, and a hierarchy was established to determine the degree of over or undersampling among these strata. CHIS 2017-2018 data were used to build the models. All available auxiliary data from voter registration databases, consumer databases, Marketing Systems Group database information (specifically, all ranges of surnames), and Census Planning Database data were appended to the CHIS 2017-2018 data. All these

appended data served as the independent variables (features) in random forest models, while self-reported attributes (demographics, etc.) served as the dependent variables.

Models for CHIS 2019-2020 were specifically designed to predict the following household attributes:

1. Korean
2. Vietnamese
3. Other Asian
4. Hispanic or Spanish-Speaker
5. Low Educational Attainment or not a US Citizen
6. Have children (under 19)

Since these six models are run independently, households can be predicted to include more than one of the six target groups. For this reason, models were applied to the sample hierarchically with preference to the higher listed model (for example, a household predicted to be Korean was scored as Korean no matter what else they might have been predicted to be).

Utilizing these models results in two additional sample groups, or strata: 1) sample records for which none of the models predicted any attribute ("Residual" sample) and 2) sample for which no auxiliary data were found ("No Match" sample). The final step in utilizing the models was to develop relative sampling fractions by which households were selected within the modeled strata.

Within each geographic and modeled stratum combination, residential addresses 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 via CATI or Web. Most frequently the child interview was completed by the randomly selected respondent who was the parent or guardian. Less frequently and only within the CATI program, an adult sufficiently knowledgeable about the child's health could complete the child interview.

Table 1-1. California county and county group strata used in the CHIS 2019-2020 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,
4. Santa Clara	24. Placer	Plumas, Sierra, Siskiyou, Trinity
5. San Bernardino	25. Merced	44. Amador, Alpine, Calaveras, Inyo,
6. Riverside	26. Butte	Mariposa, Mono, Tuolumne

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

The CHIS two-year ABS sample is of sufficient size to accomplish the second objective as well, to produce statistically stable estimates for small population groups such as racial/ethnic subgroups, children, adolescents, etc.

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 dialect), Vietnamese, Korean, and Tagalog. Tagalog was administered by phone only. 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 collaborated with UCLA on the methodology and collected data for CHIS 2019-2020, 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, one randomly selected adult in each sampled household either completed an on-line survey or was interviewed by telephone by an SSRS interviewer. In addition, the study 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. The child interview was moved in 2019-2020 to take place immediately after Section A of the adult survey and the rostering of the household. The adolescent survey took place either immediately after the adult with phone interviews or in a separate session online.

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 if the interview was completed by phone. 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. Table 1-2 shows the number of completed adult, child, and adolescent interviews in CHIS 2019-2020 by mode of interview. Note that these figures were accurate as of data collection completion for 2019-2020 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 2019-2020 interviews by mode of interview and instrument

Type of sample[1]	Adult	Child	Adolescent
Total ABS	44,109[1]	6,557	2,212
Completes by Web	40,072	6,295	2,000
Completes by phone	4,037	262	212

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

[1] Includes interviews meeting the criteria as partially complete.

Interviews in all languages were administered using SSRS's computer-assisted web interviewing and computer-assisted telephone interviewing (CAWI/CATI) system. As expected, the CATI interviews were longer in duration. The duration of the CATI interviews averaged almost 48 minutes, 26 minutes, and 23minutes for the adult, child, and adolescent interviews, respectively; the duration of the CAWI interviews averaged around 35 minutes, 13 minutes, and 17 minutes for the adult, child, and adolescent interviews, respectively. Interviews in non-English languages typically took longer to complete across

both modes: the non-English CATI interviews had an average length of about 64 minutes, 31 minutes, and 29 minutes for the adult, child, and adolescent interviews respectively; the non-English CAWI interviews had an average length of about 47 minutes, 18 minutes, and 20 minutes for the adult, child, and adolescent interviews, respectively. Just over four and half percent of the adult interviews were completed in a language other than English, as were about nine percent of all child (parent proxy) interviews and one 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 2019-2020 survey topic areas by instrument

Health status	**Adult**	**Adolescent**	**Child**
General health status	✓	✓	✓
Days missed from work or school due to health problems	✓	✓	✓
Health conditions	**Adult**	**Adolescent**	**Child**
Asthma	✓	✓	✓
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			✓
Covid-19	**Adult**	**Adolescent**	**Child**
Covid testing history and effects of pandemic	✓		
Mental health	**Adult**	**Adolescent**	**Child**
Mental health status	✓	✓	
Perceived need, access and utilization of mental health services	✓	✓	
Functional impairment, stigma, three-item loneliness scale	✓		
Suicide ideation and attempts	✓	✓	
Mental health and technology	✓	✓	
Health behaviors	**Adult**	**Adolescent**	**Child**
Dietary and nutritional intake, breastfeeding (younger than 3 years)	✓	✓	✓
Physical activity and exercise, sedentary time		✓	✓
Commute from school to home		✓	✓
Alcohol use/abuse		✓	
Cigarette and E-cigarette use	✓	✓	
Marijuana use	✓	✓	
Opioid use	✓		
Chewing tobacco, tobacco flavors	✓		
Exposure to second-hand smoke	✓		
Sexual behaviors	✓	✓	
HIV testing, HIV prevention medication (PrEP/Truvada)	✓	✓	
Contraceptive use, birth control	✓	✓	
Sexual violence	**Adult**	**Adolescent**	**Child**
Past unwanted sexual encounter	✓		

(continued)

Table 1-3. CHIS 2019-2020 survey topic areas by instrument (continued)

Women's health	**Adult**	**Adolescent**	**Child**
Pregnancy status/plans and birth control	✓	✓	
Dental health	**Adult**	**Adolescent**	**Child**
Last dental visit, main reason haven't visited dentist	✓	✓	✓
Delays in getting care			✓
Current dental insurance coverage	✓		✓
Condition of teeth	✓	✓	
Neighborhood and housing	**Adult**	**Adolescent**	**Child**
Safety, social cohesion	✓	✓	✓
Homeownership	✓		
Park use, park and neighborhood safety		✓	✓
Civic engagement, community involvement	✓	✓	
Access to and use of health care	**Adult**	**Adolescent**	**Child**
Usual source of care, visits to medical doctor	✓	✓	✓
Emergency room visits	✓	✓	✓
Delays in getting care (prescriptions and medical care)	✓	✓	✓
Communication problems with doctor	✓		✓
Timely appointment	✓	✓	✓
Access to specialist and general doctors	✓		
Tele-medical care	✓		
Care coordination	✓	✓	✓
Voter engagement	**Adult**	**Adolescent**	**Child**
Voter engagement	✓		
Food environment	**Adult**	**Adolescent**	**Child**
Access to affordable foods	✓		
Availability of food in household over past 12 months	✓		
Hunger	✓		
Health insurance	**Adult**	**Adolescent**	**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	✓	✓	✓
Medical debt, hospitalizations	✓		

(continued)

Table 1-3. CHIS 2019-2020 survey topic areas by instrument (continued)

Public program eligibility	**Adult**	**Adolescent**	**Child**
Program participation (CalWORKs, Food Stamps, SSI, SSDI, WIC, TANF)	✓	✓	✓
Assets, child support, Social security/pension, worker's compensation	✓		
Medi-Cal renewal	✓		
Reason for Medi-Cal non-participation	✓	✓	✓
Parental involvement/adult supervision	**Adult**	**Adolescent**	**Child**
Parental involvement			✓
Child care and school	**Adult**	**Adolescent**	**Child**
Current child care arrangements			✓
Paid child care	✓		
First 5 California: Talk, Read, Sing Program / Kit for New Parents			✓
Preschool/school attendance, school name		✓	✓
Caregiving	**Adult**	**Adolescent**	**Child**
Caregiving	✓		
Employment	**Adult**	**Adolescent**	**Child**
Employment status, spouse's employment status	✓		
Hours worked at all jobs	✓		
Industry and occupation, firm size	✓		
Income	**Adult**	**Adolescent**	**Child**
Respondent's and spouse's earnings last month before taxes	✓		
Household income, number of persons supported by household income	✓		
Respondent characteristics	**Adult**	**Adolescent**	**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, 2019-2020 California Health Interview Survey.

1.5 Response Rates

The overall response rates for CHIS 2019-2020 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 2019-2020, the overall household response rate was 12.2 percent (the product of the screener response rate of 16.2 percent and the extended interview response rate at the household level of 75.2 percent). CHIS uses the RR4 type response rate described in the AAPOR (The American Association for Public Opinion Research), 2016 guidelines (see more detailed in *CHIS 2019-2020 Methodology Series: Report 4 – Response Rates*).

The extended interview response rate for the ABS sample varied across the adult (72.0 percent), child (85.7 percent) and adolescent (33.2 percent) interviews. The adolescent rate includes the process of obtaining permission from a parent or guardian.

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 2019-2020 (see Table 1-4b).

Table 1-4a. CHIS response rates - Conditional

Type of Sample	Screener	Household (given screened)	Adult (given screened)	Child (given screened & eligibility)	Adolescent (given screened & permission)
Overall	16.2%	75.2%	72.0%	85.7%	33.2%

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

Table 1-4b. CHIS response rates- Unconditional

Type of Sample	Screener	Household (given screened)	Adult (given screened)	Child (given screened & eligibility)	Adolescent (given screened & permission)
Overall	16.2%	12.2%	11.6%	13.9%	5.4%

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

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

Proxy interviews were conducted for any adult who was unable to complete the extended adult interview for themselves, in order to avoid biases for health estimates of chronically ill or handicapped people. Eligible selected persons were re-contacted and offered a proxy option. In CHIS 2019-2020, either a spouse/partner or adult child completed a proxy interview for eight 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 2019-2020 accomplish the following objectives:

- Compensate for differential probabilities of selection for addresses (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 frame 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 address) and several 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 2019-2020 were created primarily from the California Department of Finance's (DOF) 2019 and 2020 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 and Human Services Agency (HHSA) region in San Diego County), and education. One limitation of using 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 DOF control totals prior to calibration.

The DOF control totals used to create the CHIS 2019-2020 weights are based on 2010 Census counts, as were those used for the 2017-2018 cycle. Please pay close attention when comparing estimates using CHIS 2019-2020 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 HSSA region, 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 2019-2020 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. SAMPLING FRAMES AND METHODS

2.1. Addressed-Based Sampling Frame

The sample design for CHIS 2019-2020 can be summarized as a stratified address-based sampling (ABS) design with strategic oversamples of households predicted to have certain attributes. This strategic oversampling is discussed in the Section 2.2. The strata are consistent with prior years, and are defined by county with sub-county substrata for Los Angeles and San Diego counties, as summarized in Table 3-2.

Since 2007, the CHIS has utilized both landline and cellphone frames. However, due to a precipitous drop in telephone response rates, and an accompanying increase in costs, UCLA and SSRS conducted extensive pilot testing in 2018 to explore the possibility of using ABS sample for future CHIS waves (Wells et al., 2018, 2019). Based on encouraging results from these preliminary pilot tests, CHIS transitioned to ABS in 2019.

The ABS sample is generated from the U.S. Postal Service's Computerized Delivery Sequence File (CDS) which includes all delivery point addresses services by the USPS. CDS provides near-complete coverage of the household population in the United States (e.g., Iannacchione, 2011; Shook-Sa, 2014; Harter et al., 2016). When drawing sample for CHIS, only records flagged as residential or mostly residential are included, as well as P.O. boxes defined as the only way a household can get mail (OWGM, that is, the homeowner has requested no mail delivery at the actual household, just the P.O. Box). Excluded are other P.O. boxes, along with seasonal and vacant households. The study does not cover institutionalized residences/group quarters (e.g., prisons, psychiatric hospitals, long-time care facilities, etc.). The ABS sample for CHIS 2019-2020 was selected via probability sampling methods, and supplied by Marketing Systems Group (MSG).

2.2. Targeting Demographics through Predictive Modeling

A recent advance in survey sampling is the use of Big Data to build predictive models of household attributes such as demographics, spoken languages, and even attitudinal metrics (Djangali et al., 2019; Dutwin, 2020; McPhee et al., 2019). The process begins by appending auxiliary data to prior survey data and using this information to build models that predict self-reported survey outcomes from auxiliary data. Future samples are then scored with the outcomes of those models, enabling the creation of strata that can be used to effectively target of specific groups.

For CHIS 2019-2020, CHIS 2017-2018 data was used to build the models. All available auxiliary data from voter registration databases, consumer databases, Marketing Systems Group database information (specifically, all ranges of surnames), and Census Planning Database data were appended to the CHIS 2017-2018 data. All these appended data served as the independent variables (features) in random forest models, while self-reported attributes (demographics, etc.) served as the dependent variables. The CHIS 2018 Fall web experiment (Wells et al., 2019) was undertaken to evaluate the efficacy of the models on secondary data. The results were reported to UCLA and the CHIS Sample Design and Survey Methodology Technical Advisory Committee and were deemed satisfactory for production-use in the 2019-2020 CHIS.

Models for the 2019-2020 CHIS were specifically designed to predict the following household attributes:

1. Korean
2. Vietnamese
3. Other Asian
4. Hispanic or Spanish-Speaker
5. Low Educational Attainment or not a US Citizen
6. Have children (under 19)

Since these six models are run independently, households can be predicted to include more than one of the six target groups. For this reason, models were applied to the sample hierarchically with preference to the higher listed model (for example, a household predicted to be Korean was scored as Korean no matter what else they might have been predicted to be).

Utilizing these models results in two additional sample groups, or strata: 1) sample records for which none of the models predicted any attribute (“Residual” sample) and 2) sample for which no auxiliary data was found (“No Match” sample). The final step in utilizing the models was to develop relative sampling fractions by which households will be selected within the modeled strata. The relative sampling fractions are a way of over- or undersampling strata in relation to each other. For the CHIS 2019-2020 model, these were determined with the aim of maximizing the different key subgroup incidences while minimizing the design effect.

The sampling fractions were:

1. Korean (1.7)
2. Vietnamese (1.34)
3. Other Asian (1.0)
4. Spanish (1.3)
5. Low Educational Attainment or Not a US Citizen (1.0)
6. Have children (2.5)
7. Residual (0.95)
8. No Match (0.3)

The process for sampling for CHIS 2019-2020 was to first generate sample 2.5 times larger than would be required in a simple random sample. Then, random subsamples within modelled strata were drawn to achieve the desired relative sampling fractions. Sampling was done this way because the modelling is a post-generation process[2], and sample must therefore be generated at the rate of the highest sampling fraction. This means that all households predicted to have children were selected for the study (since it has the highest relative sampling fraction), but only a fraction of records from other modeled strata were sampled.

2.3. Geographic Area Oversampling

The CHIS design regularly includes additional sample for specialized analyses of certain geographic areas. As has been the case in prior years, San Diego County chose to oversample for additional statistical power in CHIS 2019-2020. The oversample targeted specific overall quotas by Health and Human Services Agency (HHSA) regions, for a total of 224 additional interviews in each region.

[2] This means that the auxiliary data used to run the models does not exist on the sample frame itself and must be appended once sample is drawn.

3. SAMPLING HOUSEHOLDS

In this chapter, we describe the random sampling methodology for the CHIS design. Section 3.1 contains a description of the CHIS population of interest (also referred to as a target population), along with those who were not eligible for the study. This information provides a link between the CHIS estimates and the inferential population within California. Details of the general sampling design used to select the CHIS households is contained in Section 3.3. Here, we provide an overview of the design, followed by details on supplemental samples needed to enhance analytic capabilities for certain domains. Tables are included to identify the targeted number of completed adult interviews by strata. Section 3.4 contains information on the size of the samples selected to achieve the targets and on procedures for sample release to maintain efficiency.

3.1 Population of Interest

Estimates from CHIS represent the non-institutionalized population in California including adults (ages 18 years and older), children (ages 11 and younger), and adolescents (ages 12-17 years) living in residential households (i.e., non-group quarters). Residential households were randomly chosen through an ABS frame. Eligible residences include households, apartments, and mobile homes containing individuals with (multiple or) extended families or unrelated persons if they number less than nine. Households and persons not eligible for the CHIS include

- addresses outside the state of California;
- institutionalized residences (e.g., prisons, jails, juvenile detention facilities, psychiatric hospitals, extended-stay treatment programs, and long-time care); and
- group quarters (those with nine or more unrelated persons).

3.2 Analytic Objectives

The goal of CHIS is to provide the user community with data that will produce unbiased estimates with high precision of health and health-related metrics within each design stratum (county or groups of small counties) for adults residing in California overall and by racial/ethnic groups. We summarize the sample size for key groups to meet the analytic objectives for CHIS 2019-2020 in Table 3-1.

Overall, CHIS 2019-2020 was originally designed to yield 40,000 completed adult interviews. The targeted number of adolescent and child (proxy) interviews were established per projections from

CHIS 2017-2018 combined with the 2018 Fall web experiment and sample modeling. Targets by design strata and for the supplemental samples are discussed in detail in Section 3.3.

Table 3-1. Initial targeted number of interviews by sample characteristics

Characteristics	Interviews (n)
State-wide, Main Study, Overall	
Adults (overall)	40,000
Adolescents (overall)[1]	1,350
Children (overall)[1]	4,900
Supplemental geographic samples (adults only):	
San Diego County	1,344

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

[1] Adolescent and child targets were projected based on prior rounds of CHIS.

3.3 Sample Design

As noted earlier, the sample design for CHIS 2019-2020 can be summarized as a stratified ABS design with strategic oversamples of households predicted to have certain attributes. While CHIS 2019-2020 is the first wave of CHIS utilizing an ABS sample, much of the sample design is consistent with past iterations. Namely, CHIS 2019-2020 had a stratified design where study-eligible households were contacted and one adult resident of the household (18 years of age or older) was randomly chosen to be interviewed. Additionally, if the randomly chosen adult was the biological parent or legal guardian of a child (0-11 years of age) or an adolescent (12-17 years of age), then additional subsampling occurred for those less than 18 years of age.

Similar to the previous wave, CHIS 2019-2020 utilized 44 primary geographic strata as well as 8 Los Angeles County-specific and 6 San Diego County-specific substrata. The geographic strata are shown in Table 3-2. In addition, there are 26 Los Angeles Health Districts nested within the 8 Los Angeles-specific substrata, and CHIS 2019-2020 aimed to conduct a minimum of 200 interviews per Health District in each Health District to assess feasibility of smaller geographic stratification. These Health Districts are identified in Table 3-3. Similarly, CHIS 2019-2020 also sought to conduct a minimum of 40 interviews in each component county of multi-county stratum. These counties are identified in Table 3-4.

Targeted number of adult interviews by design strata along with relative population size in California are shown in Table 3-5. Next, we provide details on supplemental sampling.

Table 3-2. Geographic strata and sub-areas

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

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

[1] Service Planning Areas (SPAs) are analytically important substrata of Los Angeles County.

[2] Health and Human Service Agency (HHSA) regions are analytically important substrata of San Diego County.

Table 3-3. Los Angeles County Health Districts

SPA 1 – Antelope Valley
Antelope Valley
SPA 2 – San Fernando Valley
East Valley
Glendale
San Fernando
West Valley
SPA 3 – San Gabriel Valley
Alhambra
El Monte
Foothill
Pasadena
Pomona
SPA 4 – Metro
Central
Hollywood-Wilshire
Northeast
SPA 5 – West
West
SPA 6 – South
Compton
South
Southeast
Southwest
SPA 7 – East
Bellflower
East LA
San Antonio
Whittier
SPA 8 – South Bay
Harbor
Inglewood
Long Beach
Torrance

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

Table 3-4. Multi-county strata

42 - Tehama, etc.
42.1 - Tehama
42.2 - Glenn
42.3 - Colusa
43 - Del Norte, etc.
43.1 - Del Norte
43.2 - Siskiyou
43.3 - Lassen
43.4 - Trinity
43.5 - Modoc
43.6 - Plumas
43.7 - Sierra
44 - Tuolumne, etc.
44.1 - Tuolumne
44.2 - Calaveras
44.3 - Amador
44.4 - Inyo
44.5 - Mariposa
44.6 - Mono
44.7 - Alpine

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

3.3.1 Supplemental Samples for San Diego County

As with the CHIS 2017-2018, San Diego County was oversampled to collect an additional 224 interviews in each of its six Health and Human Services Agency (HHSA) regions. This resulted in a target of 724 completes in each HHSA region, for a total target of 4,344 completes in San Diego. In the tables below, we show targeted number of adult interviews by geographic strata along with relative population size in California.

The revised 2019-2020 adult interview targets including the San Diego County oversample are shown in Table 3-5.

Table 3-5. Initial and final 2019-2020 targets for completed adult interviews by geographic strata (excluding supplemental samples)

Stratum	Initial Total [1,2]	Oversamples	Final Total [1,2]	Population size [3]
State Total	40,000		41,352	
1 Los Angeles (total)[1]	7,920		7,920	Over 10 million
1.1 – Antelope Valley	500		500	
1.2 – San Fernando Valley	1,670		1,670	
1.3 – San Gabriel Valley	1,372		1,372	
1.4 – Metro	878		878	
1.5 – West	500		500	
1.6 – South	788		788	
1.7 – East	1,016		1,016	
1.8 – South Bay	1,198		1,198	
2 San Diego (total)[2]	3,000	1344	4,344	3.2 million or
2.1 – North Coastal	500	224	724	greater
2.2 – North Central	500	224	724	
2.3 – Central	500	224	724	
2.4 – South	500	224	724	
2.5 – East	500	224	724	
2.6 – North Inland	500	224	724	
3 Orange	2,460		2,460	900,000 to
4 Santa Clara	1,523		3	3.2 million
5 San Bernardino	1,528		1,528	
6 Riverside	1,732		1,732	
7 Alameda	1,369		1,369	
8 Sacramento	1,274		1,274	
9 Contra Costa	930		930	
10 Fresno	731		731	
11 San Francisco	873		873	600,000 to
12 Ventura	649		649	900,000
13 San Mateo	629		629	
14 Kern	641		641	
15 San Joaquin	541		541	
16 Sonoma	500		500	Medium
17 Stanislaus	500		500	counties
18 Santa Barbara	500		500	100,000 to
19 Solano	500		500	600,000
20 Tulare	500		500	
21 Santa Cruz	500		500	
22 Marin	500		500	

(continued)

Table 3-5. Initial and final 2019-2020 targets for completed adult interviews by geographic strata (excluding supplemental samples) (continued)

Stratum	Initial Total [1,2]	Oversamples	Final Total [1,2]	Population size [3]
23 San Luis Obispo	500		500	
24 Placer	500		500	
25 Merced	500		500	
26 Butte	500		500	
27 Shasta	500		500	
28 Yolo	500		500	
29 El Dorado	500		500	
30 Imperial	500		500	
31 Napa	500		500	
32 Kings	500		500	
33 Madera	500		500	
34 Monterey	500		500	
35 Humboldt	500		500	
36 Nevada	500		500	Small counties
37 Mendocino	500		500	Less than
38 Sutter	500		500	100,000
39 Yuba	500		500	
40 Lake	500		500	
41 San Benito	500		500	
42 Tehama, etc.	400		400	Small counties
43 Del Norte, etc.	400		400	combined
44 Tuolumne, etc.	400		400	

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

[1] Service Planning Areas (SPAs) are analytically important substrata of Los Angeles County. Counts are rounded target allocations; the sum across and by SPA differ from the total targets due to rounding.

[2] Health and Human Service Agency (HHSA) regions are analytically important substrata of San Diego County. Counts are rounded target allocations; the sum across and by HHSA region differ from the total targets due to rounding.

[3] Based on 2020 California Department of Finance population estimates excluding group quarters.

3.4 Sample Selection and Sample Releases

To meet the targets for the adult interviews outlined above, a stratified sample was selected based on the final modeled strata ratios. Where available, phone numbers were appended to the ABS sample to enable follow up protocols for nonresponse. Table 3-6 contains the total numbers of addresses randomly generated and fielded by modeled strata, and it also enumerates the number of phone appends by modeled strata. Yields were based on the Fall web experiment and adjusted to account for the design of CHIS 2019-2020.

Table 3-6. Total sample generated and fielded by modeled strata by year

Stratum	2019			2020		
Stratum	Modeled Households	Sample Mailed	Mailed Sample with Phone Appended	Modeled Households	Sample Mailed	Mailed Sample with Phone Appended
1 Korean Household	7,181	3,501	2,271	3,746	2,427	1,588
2 Vietnamese Household	9,018	3,627	2,605	5,606	2,918	2,175
3 Other Asian Household	36,273	10,807	7,382	21,266	8,117	5,772
4 Hispanic or Spanish-Speaking Household	99,324	39,385	27,186	55,425	27,460	19,776
5 Household with adult with low educational attainment or an adult who is not a citizen	57,267	17,848	11,757	36,607	14,001	9,216
6 Household with a child (under 19)	60,198	44,530	31,113	33,153	31,594	25,040
7 Residual group	422,844	120,819	82,134	248,072	90,535	65,916
8 No auxiliary data group	155,825	14,279	8,063	113,884	13,330	5,139
Total	847,930	254,796	172,511	517,759	190,382	134,622

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

All addresses sampled for CHIS 2019-2020 were checked against known or listed addresses from the Fall 2018 web experiment. Any duplicate addresses from the Fall 2018 web experiment were removed

from the CHIS 2019-2020 sample prior to data collection. Further each generation of sample in CHIS 2019-2020 was de-duped with prior releases to ensure that addresses were not duplicated.

The sample for CHIS 2019 was generated monthly and released in 9 waves. The sample for CHIS 2020 was similarly generated monthly and released in 17 waves. Tables 3-7a and b below shows the sample size and mailing date for each of the 2019 and 2020 waves.

Table 3-7a. CHIS 2019 Sample release by wave

Mail Wave	Sample Size	Initial Mailing
1	38,267	9/26/2019
2	38,281	10/3/2019
3	38,277	10/10/2019
4	27,358	10/17/2019
5	26,577	10/24/2019
6	26,576	10/31/2019
7	17,738	11/7/2019
8	35,763	11/21/2019
9	5,959	12/4/2019

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

Table 3-7b. CHIS 2020 Sample release by wave

Mail Wave	Sample Size	Initial Mailing
10	11,434	3/5/2020
11	11,431	3/12/2020
12	19,397	4/16/2020
13	19,396	4/27/2020
14	14,997	5/4/2020
15	19,994	5/7/2020
16	19,997	5/14/2020
17	14,926	5/28/2020
18	14,926	6/4/2020
19	14,926	6/18/2020
20	3,112	7/2/2020
21	3,113	7/9/2020
22	3,112	7/16/2020
23	3,854	7/30/2020
24	3,032	8/6/2020
25	3,031	8/13/2020
26	9,704	8/27/2020

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

4. WITHIN-HOUSEHOLD SAMPLING

In this chapter, we describe the random sampling methodology for the second stage of selection in the CHIS design—persons within household. One adult was randomly chosen from each household. If the selected adult was the parent of at least one child less than the age of 12, then a proxy interview was conducted for one randomly chosen child. If the selected adult was a parent of at least one adolescent (age 12-17), then an interview was conducted with a randomly chosen adolescent after receiving parent permission.

Section 4.1 contains details of the sampling design to select one adult from each eligible CHIS household. Section 4.2 contains a description of the procedures implemented to boost child (proxy) interviews in CHIS 2019-2020, and also describes the differential sampling across two child age groups. Section 4.3 provides a discussion of procedures for choosing one adolescent for interview.

4.1 Adult Sampling

As with previous waves of CHIS, adults are considered to be any person 18 years of age or older. Adult selection follows the next-birthday method of within household sampling that does not require enumerating all adults within a household. This method is intended to reduce screener duration and respondent burden, while giving each adult resident an equal probability of selection. The total number of adults in the household is collected in the screener. With this information in hand, the procedure works as follow:

- The program asks the screener respondent for the number of adults in the household.
- If only one adult lives in the household, then that adult is selected for CHIS.
- If two or more adults live in the household, respondents are asked whether they are the person with the next birthday. If so, they are chosen as the adult respondent. If not, the web program informs the respondent that the adult with the next birthday needs to complete that portion of the survey[3]. In CATI, the interviewer asks the screener respondent for the first name or initials of the adult in the household with the next birthday, and then requests to speak with that person.

[3] The verification question was adapted from Olson & Smyth (2017) to help improve selection accuracy by providing the respondent an active task. CHIS ABS pilot tests experimentally tested the verification question against alternative within-household selection approaches and found it had significantly improved selection accuracy (Wells et al., 2018, 2019).

- In CATI, if the respondent does not know who the person with the next birthday is or refuses to answer the question, the interviewer asks for the first name, age, and gender of all the adults in the household. The CATI system then randomly selects one of those adults to be the adult respondent.

4.2 Child Sampling

A child is defined for CHIS as a person less than 12 years of age normally residing in the eligible household. Eligible children are those who are the legal child of the sampled adult; foster children, or those under the informal care of a relative, are excluded from this definition. One child was selected from the eligible set rostered either in the screener under the "child-first" procedure (Section 4.2.1) or at the end of Section A of the adult questionnaire (Section 4.2.2).

As with previous CHIS waves, children 0-5 years of the selected adult were sampled at twice the rate as older children 6-11 years to increase their representation in the sample. The probability of selecting a child in the 0-5 year group was defined as $2n_1/(2n_1 + n_2)$, where n_1 was the number of eligible children ages 0-5 years and $n_2$was the number of children ages 6-11 years within the household. The corresponding selection probability for eligible children ages 6-11 years was $n_2/(2n_1 + n_2)$. Either the screener respondent or the sampled adult completed the "child interview" about the sampled child. Table 4-1 shows the distribution of households by child age category for CHIS 2019-2020.

Table 4-1. Distribution of households with children by child selection probability and year

Child selection probability	Age category of children in household[1]	2019		2020		2019-2020	
		n	pct	n	pct	n	pct
Equal	Only children 0 to 5 years	1,290	37%	1,351	31%	2,641	33%
	Only children 6 to 11 years	1,603	46%	2,008	45%	3,611	46%
Unequal	Children 0 to 5 and 6 to 11 years	611	17%	1,056	24%	1,667	21%
	Total	3,504	100%	4,415	100%	7,919	100%

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

Note: n = sample size; pct = unweighted percent.

[1] Includes all sampled households with eligible children regardless of the sampling frame and final response status.

4.2.1 Child-First Procedure

In previous CHIS waves, generally the adult interview was conducted before the child interview, and there were only a few interviews conducted through the "child-first" procedure. Starting in CHIS 2005, this child-first procedure referred to a method where, for the landline frame, a screener respondent who was not the selected adult respondent was allowed to complete the interview for an eligible child. Per protocol, the screener respondent had to be the parent of the child and be sufficiently knowledgeable to conduct the interview. Hence, under this procedure, the screener respondent was the spouse or partner of the selected adult chosen for a CHIS interview. Once the child interview was completed for landline households with an eligible adolescent, the screener respondent was asked to consent to the conducting of the adolescent interview. For CHIS 2019, we continued to follow this process for interviews conducted via the CATI mode. Table 4-2 shows the effect of the CATI child-first procedure on completed child and adolescent interviews.

Table 4-2. Effect of the child-first procedure on completed child and adolescent interviews

	2019	Pct[1,2]	2020	Pct[1,2]	2019-2020	Pct[1,2]
Households with children (total)	3,735	100.0	4,419	100.0	8,154	100.0
Child-first procedure	30	0.8	32	0.7	62	0.8
Child interview	15	50.0	15	46.9	30	48.4
No child interview	15	50.0	17	53.1	32	51.6
No child-first procedure	3,705	99.2	4,387	99.3	8,092	99.2
Child interview	2,994	80.8	3,533	80.5	6,527	80.7
No child interview	711	19.2	854	19.5	1,565	19.3
Households with adolescents (total)	3,053	100.0	3,697	100.0	6,750	100.0
Child-first procedure	9	0.3	4	0.1	13	0.2
Adolescent interview	6	66.7	2	50.0	8	61.5
No adolescent interview	3	33.3	2	50.0	5	38.5
No child-first procedure	3,044	99.7	3,693	99.9	6,737	99.8
Adolescent interview	841	27.6	1,363	36.9	2,204	32.7
No adolescent interview	2,203	72.4	2,330	63.1	4,533	67.3

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.
Note: pct = unweighted percent.
[1] Unweighted percent by child-first procedure (Y/N) taken with respect to total child/adolescent households.
[2] Unweighted percent by interview complete (Y/N) taken with respect to households by child-first procedure (Y/N).

4.2.2 Child-then-Adult Ordering

To maximize the child sample size in CHIS 2019-2020, the child rostering section was moved up to the end of Section A instead of Section G. If the adult respondent had an eligible child in the

household, the survey then proceeded with completing the child interview before resuming the adult interview. This protocol was a departure from previous waves where now essentially every child interview was conducted prior to the adult interview and could be considered a type of child-first procedure.

The key distinction between child-first completes and child-then-adult completes is that child-first completes were conducted with a sufficiently knowledgeable adult over CATI with someone who was not the selected adult respondent. Child-then-adult completes were with the selected adult respondent.

4.3 Adolescent Sampling

An adolescent is defined for CHIS as a person between the ages of 12 and 17 years normally residing in the sampled household. Like the child, the adolescent was eligible for the study only if they were the legal child of the selected sample adult. One adolescent was selected with equal probability, i.e., the selection probability was one over the number of eligible adolescents. The eligible adolescents were rostered either in the screener under the child-first procedure (Section 4.2.1) or at the end of in Section A of the adult questionnaire as with the selection of the eligible child (Section 4.2.2). The adolescent was interviewed as soon as parental permission and adolescent assent were obtained. Parental permission to interview an adolescent was obtained in Section G. This change is a departure from previous cycles of CHIS where permission was obtained from the selected adult respondent following the adult interview, or from a parent or legal guardian during the screener under the child-first procedure.

5. ACHIEVED SAMPLE SIZES

In this chapter, we detail the number of completed person-specific interviews by key characteristics for CHIS 2019-2020. Targets were set for the number of adult interviews by geographic stratum (discussed below). The relationship between the targets and achieved numbers is summarized. The associated response rates are presented in *CHIS 2019-2020 Methodology Series: Report 4 – Response Rates*. Table 5-1 compares the number of completed interviews by interview type.

Table 5-1. Number of completed interviews by type of sample and year

Interview type	Completed interviews by year								
	2019[1]	Target	Pct[2]	2020[1]	Target	Pct[2]	2019-2020[1]	Target	Pct[2]
Adult	22,160	20,676	107.2%	21,949	20,676	106.2%	44,109	41,352	106.7%
Child	3,009			3,548			6,557		
Adolescent	847			1,365			2,212		

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

Note: n = sample size; pct = unweighted percent

[1] Includes interviews meeting the criteria as partially complete.

[2] Unweighted percent is calculated as the number of completed adult interviews divided by the target within sample type.

Table 5-2 provides the distribution of completed adult interviews by geographic stratum.

Table 5-2. Number of completed adult interviews by self-reported stratum

	CHIS 2019	CHIS 2020	CHIS 2019-2020		
Reported stratum	Completes	Completes	Completes	Target	% of target
State-wide	22,160	21,949	44,109	41,352	106.67
1 Los Angeles	4,241	4,314	8,555	7,920	108.02
2 San Diego	2,443	2,297	4,740	4,344	109.12
3 Orange	1,260	1,253	2,513	2,460	102.15
4 Santa Clara	777	797	1,574	1,524	103.28
5 San Bernardino	754	839	1,593	1,528	104.25
6 Riverside	967	850	1,817	1,732	104.91
7 Alameda	681	738	1,419	1,370	103.58
8 Sacramento	645	656	1,301	1,274	102.12
9 Contra Costa	482	468	950	930	102.15
10 Fresno	438	360	798	732	109.02
11 San Francisco	412	511	923	874	105.61
12 Ventura	304	362	666	650	102.46
13 San Mateo	329	323	652	630	103.49
14 Kern	346	341	687	642	107.01
15 San Joaquin	308	283	591	542	109.04
16 Sonoma	305	276	581	500	116.20
17 Stanislaus	307	235	542	500	108.40
18 Santa Barbara	276	258	534	500	106.80
19 Solano	321	254	575	500	115.00
20 Tulare	221	282	503	500	100.60
21 Santa Cruz	240	301	541	500	108.20
22 Marin	287	242	529	500	105.80

(continued)

Table *5-2*. Number of completed adult interviews by self-reported stratum (continued)

	CHIS 2019	CHIS 2020	CHIS 2019-2020		
Reported stratum	Completes	Completes	Completes	Target	*% of target*
23 San Luis Obispo	227	289	516	500	103.20
24 Placer	256	284	540	500	108.00
25 Merced	229	290	519	500	103.80
26 Butte	257	260	517	500	103.40
27 Shasta	248	286	534	500	106.80
28 Yolo	245	285	530	500	106.00
29 El Dorado	268	238	506	500	101.20
30 Imperial	275	249	524	500	104.80
31 Napa	266	293	559	500	111.80
32 Kings	323	251	574	500	114.80
33 Madera	252	273	525	500	105.00
34 Monterey	244	261	505	500	101.00
35 Humboldt	289	259	548	500	109.60
36 Nevada	247	272	519	500	103.80
37 Mendocino	250	259	509	500	101.80
38 Sutter	286	264	550	500	110.00
39 Yuba	271	235	506	500	101.20
40 Lake	278	240	518	500	103.60
41 San Benito	224	274	498	500	99.60
42 Tehama, etc	277	216	493	400	123.25
43 Del Norte, etc.	282	212	494	400	123.50
44 Tuolumne, etc.	322	219	541	400	135.25

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

[1] Includes interviews meeting the criteria as partially complete.

Table 5-3 and Table 5-4 contain the number of completed child and adolescent interviews distributed by reported geographic stratum.

Table 5-3. Number of completed child interviews by self-reported stratum

	CHIS 2019	CHIS 2020	CHIS 2019-2020
Reported stratum	Total	Total	Total
State-wide	3,009	3,548	6,557
1 Los Angeles	577	689	1,266
2 San Diego	341	391	732
3 Orange	164	175	339
4 Santa Clara	132	154	286
5 San Bernardino	111	155	266
6 Riverside	136	148	284
7 Alameda	92	110	202
8 Sacramento	82	111	193
9 Contra Costa	50	81	131
10 Fresno	74	70	144
11 San Francisco	50	69	119
12 Ventura	34	59	93
13 San Mateo	48	50	98
14 Kern	66	79	145
15 San Joaquin	46	51	97
16 Sonoma	31	40	71
17 Stanislaus	48	43	91
18 Santa Barbara	30	38	68
19 Solano	45	22	67
20 Tulare	45	56	101
21 Santa Cruz	18	31	49
22 Marin	34	41	75
23 San Luis Obispo	26	49	75
24 Placer	35	54	89
25 Merced	46	56	102
26 Butte	25	46	71
27 Shasta	28	32	60
28 Yolo	37	51	88
29 El Dorado	30	27	57
30 Imperial	64	48	112
31 Napa	23	40	63
32 Kings	60	58	118
33 Madera	50	42	92
34 Monterey	29	39	68
35 Humboldt	23	33	56
36 Nevada	19	31	50
37 Mendocino	17	30	47
38 Sutter	48	43	91

(continued)

Table 5-3. Number of completed child interviews by self-reported stratum (continued)

Reported stratum	CHIS 2019 Total	CHIS 2020 Total	CHIS 2019-2020 Total
39 Yuba	44	53	97
40 Lake	22	35	57
41 San Benito	29	39	68
42 Tehama, etc.	42	34	76
43 Del Norte, etc.	38	23	61
44 Tuolumne, etc.	20	22	42

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

Table 5-4. Number of completed adolescent interviews by self-reported geographic stratum

Reported stratum	CHIS 2019 Total	CHIS 2020 Total	CHIS 2019-2020 Total
State-wide	847	1,365	2,212
1 Los Angeles	163	219	382
2 San Diego	79	172	251
3 Orange	58	81	139
4 Santa Clara	31	54	85
5 San Bernardino	35	60	95
6 Riverside	43	68	111
7 Alameda	29	44	73
8 Sacramento	29	28	57
9 Contra Costa	11	25	36
10 Fresno	25	31	56
11 San Francisco	14	21	35
12 Ventura	11	20	31
13 San Mateo	17	18	35
14 Kern	12	20	32
15 San Joaquin	18	22	40
16 Sonoma	9	9	18
17 Stanislaus	8	18	26
18 Santa Barbara	7	18	25
19 Solano	5	5	10
20 Tulare	11	20	31
21 Santa Cruz	9	16	25
22 Marin	14	22	36
23 San Luis Obispo	1	15	16
24 Placer	14	20	34
25 Merced	11	22	33
26 Butte	5	11	16
27 Shasta	5	21	26
28 Yolo	16	24	40
29 El Dorado	10	18	28
30 Imperial	14	21	35
31 Napa	6	16	22
32 Kings	19	26	45
33 Madera	7	23	30
34 Monterey	11	16	27
35 Humboldt	12	16	28
36 Nevada	2	11	13
37 Mendocino	7	11	18
38 Sutter	9	17	26
39 Yuba	14	21	35
40 Lake	7	16	23
41 San Benito	9	21	30
42 Tehama, etc.	12	11	23
43 Del Norte, etc.	8	8	16
44 Tuolumne, etc.	10	9	19

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

Table 5-5 contains the number of completed interviews by family structure. As shown, we accepted proxy interviews for children and adolescent interviews for households without a corresponding adult interview. A child or adolescent interview could be completed without a corresponding adult interview under the child first procedure in CATI. Further, starting with CHIS 2019-2020, the child and adolescent rostering were placed after Section A of the adult interview. If there was an eligible child in the household, the child interview was completed before resuming the adult interview. If there was an eligible adolescent in the household, the adolescent interview could be completed after parental permission (for a detailed discussion please refer to *Section 4.2* above).

Table 5-5. Number of completed interviews by interview combinations and year

Interview combinations[1]	CHIS 2019		CHIS 2020		CHIS 2019-2020	
	n	pct	n	pct	n	pct
Adult only	19,178	84.3	18,207	80.3	37,385	82.3
Adult and child	2,195	9.7	2,477	10.9	4,672	10.3
Adult and adolescent	511	2.2	844	3.7	1,355	3.0
Adult, child, and adolescent	276	1.2	421	1.9	697	1.5
Child only	518	2.3	612	2.7	1,130	2.5
Adolescent only	40	0.2	62	0.3	102	0.2
Child and adolescent only	20	0.1	38	0.2	58	0.1
Total	22,738	100.0	22,661	100.0	45,399	100.0

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.
Note: n = sample size; pct = unweighted percent.
[1] Includes completed and partial interviews.

6. EVALUATING THE PREDICTIVE MODELING

The purpose of sampling by modeled strata in addition to geographic strata was to better target specific, difficult-to-reach groups of interest, including Koreans, Spanish speakers, and adults with children. The predictive models were very effective in targeting some groups, but less effective with others.

Table 6-1 shows the efficacy of the individual modeled predictors, before applying the hierarchical stratification, in reaching the target groups. This means the numbers in the columns do not add up to the total number of interviews shown in the last row, since a household may be predicted to have multiple attributes. This allows for a more straightforward evaluation of the predictors, since cases in higher strata are not held out when evaluating predictors used in lower strata. Similarly, numbers in the rows do not add up to the number of completed interviews shown in the last column. This is because a completed interview could qualify for more than one of the targeted demographics (e.g., a Korean respondent with a child in the household). The modeled predictors performed best when targeting ethnic groups and languages spoken. For instance, of the adult completed interviews from the modeled likely-Korean households, 241 respondents self-identified as Korean (Table 6-1). Of the adult completed interviews not from the modeled likely-Korean households, 389 respondents self-identified as Korean. However, the total number of adult completed interviews from the modeled likely-Korean households was 613, while the total number of adult completed interviews from the households not predicted to be Korean was 43,496. Therefore, the incidence of adult respondents who self-identified as Korean was 39.3% among the households predicted to be Korean and only 0.9% among the households not predicted to be Korean (Table 6-2). Similarly, the incidence of adult respondents who self-identified as Vietnamese was 47.2% among the households predicted to be Vietnamese and only 0.5% among the households not predicted to be Vietnamese.

The targeting of adults with low educational attainment or who are not US citizens was somewhat less effective. The incidence of adult respondents in households predicted to have a member who did not complete high school or was not a citizen who self-identified as not completing high school was 7.7%, which, though higher than the incidence of the same among respondents from households not predicted to have a member who did not complete high school or was not a citizen who self-identified as low educated (2.6%), did not show the same level of improvement as the predictors targeting specific race groups.

Among households predicted to contain children, the incidence of the presence of a child or adolescent among adult respondents was 46.1%, which is an improvement over the incidence of the presence of a child or adolescent among households no predicted to contain children (18.9%). The

magnitude of the improvement was not as large as what was seen in the Korean, Vietnamese, Other Asian, and Hispanic/Spanish-Speaker modeled strata, but was better than the results from the "low educational attainment/not US citizen stratum.

Table 6-1. Complete adult interviews by modeled variables (counts)

Modeled Stratum	Self-Reported Demographics							Total Completed Adult Interviews[1]
	Korean	Vietnamese	Other Asian	Hispanic / Interview Conducted in Spanish	Less than High School	Not Citizen	Child / Adolescent Present in Household	
Likely Korean Household	241	9	133	65	22	64	143	613
Likely Vietnamese Household	17	300	184	27	44	29	213	636
Likely Other Asian Household	292	313	1,298	275	114	235	829	3,151
Likely Hispanic or Spanish-Speaking Household	25	12	251	2,215	445	300	1,299	3,930
Likely Household with adult with low educational attainment or an adult who is not a citizen	310	252	1,504	2,467	569	560	2,194	7,411
Likely Household with child (under 19)	111	83	959	2,124	290	252	4,318	9,372
Residual group	126	88	1,691	3,651	586	577	3,957	24,452
No auxiliary data group	41	25	240	434	97	146	443	2,003
Overall	630	513	4,656	8,771	1,531	1,584	10,897	44,109

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.
[1] Includes interviews meeting the criteria as partially complete.

Table 6-2. Competed adult interviews by modeled variables (percentages)

Modeled Stratum	Self-Reported Demographics							Total Completed Adult Interviews[1]
	Korean	Vietnamese	Other Asian	Hispanic / Interview Conducted in Spanish	Less than High School	Not Citizen	Child / Adolescent Present in Household	
Likely Korean Household	39.3%	1.5%	21.7%	10.6%	3.6%	10.4%	23.3%	613
Likely Vietnamese Household	2.7%	47.2%	28.9%	4.2%	6.9%	4.6%	33.5%	636
Likely Other Asian Household	9.3%	9.9%	41.2%	8.7%	3.6%	7.5%	26.3%	3,151
Likely Hispanic or Spanish-Speaking Household	0.6%	0.3%	6.4%	56.4%	11.3%	7.6%	33.1%	3,930
Likely Household with adult with low educational attainment or an adult who is not a citizen	4.2%	3.4%	20.3%	33.3%	7.7%	7.6%	29.6%	7,411
Likely Household with child (under 19)	1.2%	0.9%	10.2%	22.7%	3.1%	2.7%	46.1%	9,372
Residual group	0.5%	0.4%	6.9%	14.9%	2.4%	2.4%	16.2%	24,452
No auxiliary data group	2.0%	1.2%	12.0%	21.7%	4.8%	7.3%	22.1%	2,003
Overall	1.4%	1.2%	10.6%	19.9%	3.5%	3.6%	24.7%	44,109

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.

[1] Includes interviews meeting the criteria as partially complete.

7. REFERENCES

Djangali, A.L., Joseph-David, J., & Trofimovich, L. (2019). *Increasing representativeness through the use of predictive modeling and targeted outreach*. Paper presented at the American Association of Public Opinion Research Conference, Toronto, ON.

Dutwin, D. (2020). Feedback Loop: Using Surveys to Build and Assess Registration-Based Sample Religious Flags for Survey Research. In *Big Data Meets Survey Science: A Collection of Innovative Methods*. John Wiley & Sons.

Harter, R., Battaglia, M. P., Buskirk, T. D., Dillman, D. A., English, N., Fahimi, M., Frankel, M. R., Kennel, T., McMichael, J. P., McPhee, C. B., Montaquila, J., Yancey, T., & Zukerberg, A. L. (2016). *Address-based sampling*. Oakbrook Terrace, IL: American Association for Public Opinion Research. https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx.

Iannacchione, V. (2011). The changing role of address-based sampling in survey research. *Public Opinion Quarterly*, *75*(3), 556–575.

McPhee, C., Jackson, M. & Quenneville, G. (2019). *Using model-based stratification to improve sampling efficiency*. Paper presented at the American Association of Public Opinion Research Conference, Toronto, ON.

Olson, K. & Smyth, J. (2017). Within-household selection in mail surveys: Explicit questions are better than cover letter instructions. *Public Opinion Quarterly*, *81*(3), 688-713.

Shook-Sa, B. E. (2014). Improving the efficiency of address-based sampling frames with the USPS No-Stat File. *Survey Practice*, *7*(4). http://www.surveypractice.org/index.php/SurveyPractice/article/view/264/html_8

Wells, B. M., Hughes, T., Park, R., CHIS Redesign Working Group, Rogers, T. B., & Ponce, N. (2018). *Evaluating the California Health Interview Survey of the future: Results from a methodological experiment to test an address-based sampling mail push-to-web data collection.* Los Angeles, CA: UCLA Center for Health Policy Research.

Wells, B. M., Hughes, T., Park, R., CHIS Redesign Working Group, & Ponce, N. (2019). *Evaluating the California Health Interview Survey of the future: Results from a statewide pilot of an address-*

based sampling mail push-to-web data collection. Los Angeles, CA: UCLA Center for Health Policy Research.

APPENDIX A

Appendix A contains supplemental information on the CHIS 2019-2020 sample design.

Table A-1 compares the definitions of the design strata since CHIS 2001 through the current study.

Table A-2 provides the size of the samples for CHIS 2019-2020 separately by design stratum.

Table A-1. Design strata definitions for CHIS 2001, 2003, 2005, 2007, 2009, 2011-2012, 2013-2014, 2015-2016, 2017-2018 and 2019-2020

County	2015-2016, 2017-2018, 2019-2020 Strata	2013-2014 Strata	2005, 2007, 2009, 2011-2012 Strata	2001, 2003 Strata
Los Angeles	1	1	1	1
San Diego	2	2	2	2
Orange	3	3	3	3
Santa Clara	4	4	4	4
San Bernardino	5	5	5	5
Riverside	6	6	6	6
Alameda	7	7	7	7
Sacramento	8	8	8	8
Contra Costa	9	9	9	9
Fresno	10	10	10	10
San Francisco	11	11	11	11
Ventura	12	12	12	12
San Mateo	13	13	13	13
Kern	14	14	14	14
San Joaquin	15	15	15	15
Sonoma	16	16	16	16
Stanislaus	17	17	17	17
Santa Barbara	18	18	18	18
Solano	19	19	19	19
Tulare	20	20	20	20
Santa Cruz	21	21	21	21
Marin	22	22	22	22
San Luis Obispo	23	23	23	23
Placer	24	24	24	24
Merced	25	25	25	25
Butte	26	26	26	26
Shasta	27	27	27	27
Yolo	28	28	28	28
El Dorado	29	29	29	29
Imperial	30	30	30	30
Napa	31	31	31	31
Kings	32	32	32	32
Madera	33	33	33	33
Monterey	34	34	34	34
San Benito	41	41	41	

(continued)

Table A-1. Design strata definitions for CHIS 2001, 2003, 2005, 2007, 2009, 2011-2012, 2013-2014, 2015-2016, 2017-2018, and 2019-2020 (continued)

County	2015-2016, 2017-2018, 2019-2020 Strata	2013-2014 Strata	2005, 2007, 2009, 2011-2012 Strata	2001, 2003 Strata
Lake	40	40	40	37
Mendocino	37	37	37	
Sutter	38	38	38	39
Yuba	39	39	39	
Colusa Glenn Tehama	42	42	42	38
Nevada	36	36	36	40
Humboldt	35	35	35	35
Del Norte	43	43	43	
Lassen Modoc Plumas				36
Sierra Trinity				40
Siskiyou		43.2		36
Amador Alpine Inyo Mariposa Mono	44	44	44	41
Tuolumne		44.1		
Calaveras		44.2		

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey

Table A-2. Number of sample pieces selected by design stratum

Sampling stratum	2019 Total	2019 Sample with Appended phones	2020 Total	2020 Sample with Appended phones	2019-2020 Total	2019-2020 Sample with Appended phones
State-wide	254,796	172,511	190,382	134,622	445,178	307,133
1 Los Angeles	62,254	43,346	42,781	30,609	105,035	73,955
2 San Diego	26,785	18,235	18,305	13,165	45,090	31,400
3 Orange	13,271	9,532	9,847	7,428	23,118	16,960
4 Santa Clara	8,133	5,215	5,663	3,870	13,796	9,085
5 San Bernardino	10,327	7,254	9,262	6,825	19,589	14,079
6 Riverside	11,671	8,830	8,685	6,916	20,356	15,746
7 Alameda	6,524	4,357	5,123	3,592	11,647	7,949
8 Sacramento	6,377	4,125	4,739	3,227	11,116	7,352
9 Contra Costa	4,841	3,525	3,361	2,517	8,202	6,042
10 Fresno	5,966	3,660	3,762	2,401	9,728	6,061
11 San Francisco	4,674	2,761	3,580	2,244	8,254	5,005
12 Ventura	2,602	2,071	2,562	2,111	5,164	4,182
13 San Mateo	3,191	2,193	2,362	1,728	5,553	3,921
14 Kern	4,364	2,751	3,639	2,471	8,003	5,222
15 San Joaquin	4,311	2,975	3,245	2,325	7,556	5,300
16 Sonoma	2,622	1,791	1,859	1,338	4,481	3,129
17 Stanislaus	3,579	2,361	2,417	1,670	5,996	4,031
18 Santa Barbara	2,578	1,675	1,721	1,197	4,299	2,872
19 Solano	3,583	2,462	2,337	1,666	5,920	4,128
20 Tulare	2,712	1,711	3,072	2,000	5,784	3,711
21 Santa Cruz	2,480	1,679	1,747	1,286	4,227	2,965
22 Marin	2,425	1,716	1,524	1,123	3,949	2,839
23 San Luis Obispo	1,539	1,067	1,772	1,221	3,311	2,288
24 Placer	2,211	1,558	1,967	1,464	4,178	3,022
25 Merced	3,130	1,982	3,086	1,954	6,216	3,936
26 Butte	2,314	1,460	1,882	1,253	4,196	2,713
27 Shasta	1,932	1,310	2,237	1,573	4,169	2,883
28 Yolo	2,062	1,310	1,588	1,052	3,650	2,362
29 El Dorado	2,189	1,549	1,464	1,058	3,653	2,607
30 Imperial	4,020	2,435	2,941	1,950	6,961	4,385
31 Napa	2,634	1,870	2,270	1,661	4,904	3,531
32 Kings	4,669	2,791	3,159	2,026	7,828	4,817
33 Madera	3,581	2,218	2,865	1,810	6,446	4,028
34 Monterey	2,893	1,883	2,411	1,706	5,304	3,589
35 Humboldt	2,128	1,360	1,450	1,045	3,578	2,405
36 Nevada	1,903	1,365	1,663	1,248	3,566	2,613
37 Mendocino	2,146	1,334	1,745	1,107	3,891	2,441
38 Sutter	3,588	2,334	3,146	2,192	6,734	4,526
39 Yuba	3,446	2,122	2,624	1,658	6,070	3,780
40 Lake	2,843	1,816	2,173	1,428	5,016	3,244
41 San Benito	2,552	1,693	2,972	1,997	5,524	3,690
42 Tehama, etc.	3,316	2,063	2,211	1,471	5,527	3,534
43 Del Norte, etc.	1,976	1,151	1,644	1,002	3,620	2,153
44 Tuolumne, etc.	2,454	1,615	1,519	1,037	3,973	2,652

Source: UCLA Center for Health Policy Research, 2019-2020 California Health Interview Survey.