california
health
interview
survey

CHIS 2023 Methodology Report Series

Report 1

Sample Design

September 2024

UCLA Center for Health Policy Research

CALIFORNIA HEALTH INTERVIEW SURVEY

CHIS 2023 METHODOLOGY SERIES

REPORT 1

SAMPLE DESIGN

September 2024

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

www.chis.ucla.edu

This report provides analysts with information about the sampling methods used for CHIS 2023, 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 2023 Methodology Series: Report 1 - Sample Design*. Los Angeles, CA: UCLA Center for Health Policy Research, 2024.

The California Health Interview Survey is a collaborative project of the UCLA Center for Health Policy Research with multiple funding sources. Funding for CHIS 2023 came from the following sources: the California Department of Health Care Services, the California Department of Health Care Services (Community Services Division), the California Department of Public Health, The California Endowment, the California Health Benefit Exchange, the California Health Care Foundation, the California Tobacco Prevention Program, the California Wellness Foundation, First 5 California, the California Civil Rights Department, Santa Clara County Public Health Department, City of Long Beach Department of Health and Human Services, and San Diego County Health and Human Services Agency.

PREFACE

Sample Design is the first in a series of methodological reports describing the 2023 California Health Interview Survey (CHIS 2023). 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 2023 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 2023

The methodological reports for CHIS 2023 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 2023. An appropriate sample design is a feature of a successful survey, and CHIS 2023 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 adolesnts 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 2023 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 https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-methods/chis-methodology-reports-repository . 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

Table of Figures

1. CHIS 2023 SAMPLE DESIGN AND METHODOLOGY SUMMARY

1.1 Overview

A series of five methodology reports is available with more detail about the methods used in CHIS 2023.

- 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 https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-methods/chis-methodology-reports-repository.

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 https://healthpolicy.ucla.edu/our-work/publications for examples of CHIS studies).

1.2 Sample Additions and Data Collection Methodology Updates

Starting in 2021, the CHIS added a prepaid cell phone sample to the primary ABS sample. A second innovation was altering the envelope for the initial mailing to have a window that would allow the incentive to be seen. The CHIS research team deemed these changes necessary to improve representation of California's diverse population and improve response rates.

For CHIS 2023, respondents in the ABS sample 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, a standard letter, and a final postcard. Where addresses can be matched to a listed telephone number, the nonresponding households are also called up to six times to attempt to complete an interview before the sampled household is considered to be a resolved nonresponse. In addition to the ABS sample frame, CHIS 2023 utilized a supplemental listed prepaid cell phone sample to meet targets in certain stratum.

The prepaid cell phone oversample followed the same dialing protocol of up to six dials before retiring the sample. In addition, the sampled phone number was screened for respondents who were either aged 18 to 24, Hispanic, African American, or would take the survey in one of the non-English languages offered for CHIS 2023.

In addition to the prepaid cell phone oversample, CHIS 2023 included two geographic oversamples:

1) An oversample of households from 11 ZIP codes in the City of Long Beach.
2) An oversample of households in Santa Clara County.

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 https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/access-chis-data/resources .

Additional documentation on constructing the CHIS sampling weights is available in the *CHIS 2023 Methodology Series: Report 5—Weighting and Variance Estimation* posted at https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-methods/chis-methodology-reports-repository. 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 2023 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 2023 continued to employ an address-based sample design. For the ABS sample, the 58 counties in the state were grouped into 44 primary 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.

As with CHIS 2021-2022, the address-based sample in CHIS 2023 was stratified into different strata that had higher incidences of individuals with targeted characteristics. For CHIS 2023, these strata were based on predictive models that employed Big Data techniques to identify household attributes such as demographics, spoken languages, and even attitudinal metrics that are correlated with important respondent characteristics. The process begins by taking prior data and building models with those data, and then scoring future samples with the outcomes of those models. In addition to evaluating the predictive models, for CHIS 2023 we also investigated the utility of individual sample flags provided by MSG database information, including the surname flags, child indicator variables, and resident age information as well as PDB block-group characteristics including the density of households with African American residents and households with limited English proficiency.

For CHIS 2023, the following strata were created[1]:

1) Vietnamese
2) Korean
3) Likely Asian-language Interview
4) Likely Spanish-language interview
5) Hispanic
6) Other high-density non-English
7) Other Asian
8) High density African American
9) HH with children
10) Other 65+
11) Residual - Match
12) Residual – No match

This stratification scheme was designed to make use of the most effective predictive variables to target key demographic subgroups in an efficient way that minimizes the impact of the disproportionate sampling on the design effect. Those models that were not sufficiently predictive to add value were excluded. It should be noted that this stratification includes two additional strata: 1) sample records for which none of the variables or models predicted any attribute, but for which auxiliary data could be matched to the address ("Residual - Match" sample) and sample for which no Big Data was found ("Residual - No match" sample). The final step in utilizing the models is to develop sampling fractions by which modeled households will be selected. The final sample fractions balanced the need to increase the frequency of the lowest incidence groups, while accounting for subgroup differences in response propensity and minimizing disproportionate weighting whenever possible.

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 selected parent/guardian were randomly selected. The adolescent was interviewed directly via CATI or Web. The child interview was completed by the selected adult respondent who was the parent or guardian.

[1] The Santa Clara oversample employs a slightly different strata, please refer to Methodology Report 1 – Sample Design for additional details.

Table 1-1. California county and county group strata used in the CHIS 2023 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, 2023 California Health Interview Survey.

In addition to the ABS sample frame, CHIS 2023 utilized a supplemental listed prepaid cellphone sample to meet targets in twelve geographic stratum that were underperforming in completion rate. Listed prepaid cell phones were sampled from the following 12 geographic strata:

1. Los Angeles
    a. SPA1
    b. SPA5
2. San Diego
    a. Central
3. Santa Clara
4. Sacramento
5. Contra Costa

6. Ventura
7. San Joaquin
8. Sonoma
9. Santa Cruz
10. Merced
11. Mendocino
12. San Benito

To better target populations not adequately covered under the ABS frame in CHIS 2023, we utilized a prepaid cell phone oversample of 450 completes to obtain additional in-language interviews, Hispanic and African American samples, and young adults. Prepaid cell phone numbers are associated with cell phones that are "pay-as-you-go" and do not require a contract. Prepaid numbers are more likely to be used by Hispanics, people with lower education and lower income, and other related groups that are often underrepresented in general population samples (e.g., the uninsured)

The CHIS ABS sample and the prepaid oversample were of sufficient size to accomplish the second objective, i.e., 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. These languages were chosen based on analysis of ACS 2021 5-year 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 2023, 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 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.

Table 1-2 shows the number of completed adult, child, and adolescent interviews in CHIS 2023 by mode of interview. Note that these figures were accurate as of data collection completion for 2023 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 https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-methods/chis-design.

Table 1-2. Number of completed interviews by mode of interview and instrument

	Adult	Child	Adolescent
Totals[1]	23,697	3,650	1,045
Completes by Web	21,101	3,370	989
Completes by phone	2,596	280	56

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.
[1] Includes interviews meeting the criteria of sufficient partial.

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 68 minutes, 20 minutes, and 25 minutes for the adult, child, and adolescent interviews, respectively; the duration of the CAWI interviews averaged around 45 minutes, 13 minutes, and 18 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 76 minutes, 22 minutes, and 25 minutes for the adult, child, and adolescent interviews respectively; the non-English CAWI interviews had an average length of about 54 minutes, 16 minutes, and 18 minutes for the adult, child, and adolescent interviews, respectively.

Nearly 8 percent of the adult interviews were completed in a language other than English, as were about 12 percent of all child (parent proxy) interviews and 2 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 2023 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	✓		
Heart disease, High blood pressure, Cholesterol	✓		
Physical disability	✓		
Mental health	**Adult**	**Adolescent**	**Child**
Mental health status	✓	✓	
Perceived need, Access and utilization of mental health services	✓	✓	
Functional impairment, Stigma	✓		
Suicide ideation and attempts	✓	✓	
Telehealth and mental health services satisfaction, Delays in mental health services	✓	✓	
Climate Change	✓	✓	
Health behaviors	**Adult**	**Adolescent**	**Child**
Moderate physical activity	✓		
Dietary intake	✓	✓	✓
Breastfeeding (younger than 3 years)			✓
Sugar-sweetened beverages	✓	✓	✓
Alcohol use, Cigarette use, E-cigarette use, Marijuana use, CBD use	✓	✓	
CBD Use	✓		
Opioid use, Prescription painkiller use	✓		
Exposure to second-hand smoke/vapor, Exposure to marijuana smoke	✓		
Sexual behaviors, HIV testing, HIV prevention medication	✓	✓	
Caregiving	✓		
Gambling, Financial and mental impacts of gambling	✓		
Gun Violence	**Adult**	**Adolescent**	**Child**
Firearm ownership/presence, Loaded, and secure, Firearm victimization, Quick access to firearm	✓		
Women's health	**Adult**	**Adolescent**	**Child**
Pregnancy status	✓		

(continued)

Table 1-3. CHIS 2023 survey topic areas by instrument (continued)

Dental health	**Adult**	**Adolescent**	**Child**
Last dental visit, Main reason have not visited dentist, Number of dental visits, Location of dental service	✓	✓	✓
Current dental insurance coverage	✓	✓	✓
Source of dental care	✓	✓	✓
Neighborhood and housing	**Adult**	**Adolescent**	**Child**
Safety, Social cohesion	✓	✓	✓
Civic engagement	✓	✓	
Participation in extracurricular activities		✓	
Housing security/stability, Place of residency last year	✓		
Encounters with police	✓		
Adverse Childhood Experiences	**Adult**	**Adolescent**	**Child**
ACES Screener	✓	✓	
Past ACES screener	✓	✓	
Safe and nurtured childhood experiences	✓	✓	
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	✓		✓
Contraception	✓	✓	
Timely appointment	✓	✓	✓
Access to specialist and general doctors	✓		
Telehealth care, Telehealth visit satisfaction and barriers	✓		
Care coordination	✓	✓	✓
Discrimination in healthcare setting	✓		
Difficulty in accessing care, tests, treatment	✓	✓	✓
Voter engagement	**Adult**	**Adolescent**	**Child**
Voter engagement	✓		
Voter attitudes	✓		
Food environment	**Adult**	**Adolescent**	**Child**
Availability of food in household over past 12 months, Hunger	✓		

(continued)

Table 1-3. CHIS 2023 survey topic areas by instrument (continued)

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	✓	✓	✓
Partial scope Medi-Cal, Medical debt, Hospitalizations	✓		
Public program eligibility	**Adult**	**Adolescent**	**Child**
Household poverty level	✓		
Program participation (CalWORKs, Food Stamps, SSI, SSDI, WIC, TANF)	✓	✓	✓
Assets, Child support, Social security/pension, Worker's compensation	✓		
Medi-Cal eligibility, Medi-Cal renewal, Notice of actions from Medi-Cal	✓		
Reason for Medi-Cal non-participation among potential beneficiaries	✓	✓	✓
Use of public benefits among immigrant residents	✓		
Parental involvement/adult supervision	**Adult**	**Adolescent**	**Child**
Parental involvement			✓
Book ownership, Source of reading materials, Challenges to reading to child			✓
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		✓	✓
Preschool quality			✓
Employment	**Adult**	**Adolescent**	**Child**
Employment status, Spouse's employment status	✓		
Hours worked at all jobs	✓		
Industry and occupation, Firm size	✓		
Paid Family Leave	✓		
Income	**Adult**	**Adolescent**	**Child**
Respondent's and spouse's earnings last month before taxes	✓		
Household income, Number of persons supported by household income	✓		

(continued)

Table 1-3. CHIS 2023 survey topic areas by instrument (continued)

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	✓	✓	✓
COVID-19	**Adult**	**Adolescent**	**Child**
Ever tested positive for COVID-19, Test type	✓		
Experienced long COVID-19 symptoms	✓		
COVID vaccine status, COVID booster status	✓	✓	✓
Future COVID vaccine acceptance, Reasons for COVID vaccine hesitancy	✓		
Challenges experience due to COVID-19 pandemic	✓		
N95 masks, Ability to get N95 masks	✓		
Adolescent Future Preparedness	**Adult**	**Adolescent**	**Child**
Plans for college, Impact of pandemic on college plans		✓	
Discrimination	**Adult**	**Adolescent**	**Child**
Housing discrimination experience, Main reason for discrimination, Housing Choice Section 8 Voucher	✓		
Hate incident experience and witness, Type, Location, Reason for hate incident	✓	✓	

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.

1.5 Response Rates

The overall response rates for CHIS 2023 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 2023, the overall household response rate was 8.5 percent (the product of the screener response rate of 11.8 percent and the extended interview response rate at the household level of 72.1 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 2023 Methodology Series: Report 4 – Response Rates*).

The extended interview response rate for the ABS sample varied across the adult (64.7 percent), child (82.2 percent) and adolescent (27.9 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 2023 (see Table 1-4b).

Table 1-4a. CHIS response rates - Conditional

Type of Sample	Screener[1]	Household (given screened)[1]	Adult (given screened)[1]	Child (given screened & eligibility)[1]	Adolescent (given screened & permission)[1]
Overall	11.8%	72.1%	64.7%	82.2%	27.9%

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.

[1] The prepaid cell, Long Beach, and Santa Clara oversamples are not included in these rates.

Table 1-4b. CHIS response rates - Unconditional

Type of Sample	Screener[1]	Household (given screened)[1]	Adult (given screened)[1]	Child (given screened & eligibility)[1]	Adolescent (given screened & permission)[1]
Overall	11.8%	8.5%	7.7%	9.7%	3.3%

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.

[1] The prepaid cell, Long Beach, and Santa Clara oversamples are not included in these rates.

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 "sufficient partial 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 2023, either a spouse/partner or adult child completed a proxy interview for sixteen 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 https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-methods/chis-design/chis-2019-2020-redesign.

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 2023 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 under coverage 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 2023 were primarily created from the California Department of Finance's (DOF) 2023 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 2023 weights are based on 2020 Census counts. Please pay close attention when comparing estimates using CHIS 2023 data with estimates using data from CHIS cycles before 2023. 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 2010 Census-based information and 2020 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 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 zipcode 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 2023 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

The sample design for CHIS 2023 included both a stratified address-based sample (ABS) and a listed prepaid cell phone sample. Details of the sample design are discussed below.

2.1 Addressed-Based Sampling Frame

The sample design for CHIS 2023 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 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, 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) are included. Other P.O. boxes, along with seasonal and vacant households are excluded. The study does not cover institutionalized residences/group quarters (e.g., prisons, psychiatric hospitals, long-time care facilities, etc.). The ABS sample for CHIS 2023 was selected via probability sampling methods and supplied by Marketing Systems Group (MSG).

2.2 Listed Prepaid Cellphone Sample

In addition to the ABS sample frame, CHIS 2023 utilized a supplemental listed prepaid cellphone sample to meet targets in stratum that were underperforming in completion rate. In order to meet the target number of completes, 36,835 pieces of listed prepaid sample were sampled from the following 12 geographic strata:

13. Los Angeles
    a. SPA1
    b. SPA5
14. San Diego
    a. Central
15. Santa Clara
16. Sacramento
17. Contra Costa
18. Ventura
19. San Joaquin
20. Sonoma
21. Santa Cruz
22. Merced
23. Mendocino
24. San Benito

2.3. Targeting Demographics through Predictive Modeling

An 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 specific groups.

For CHIS 2023, SSRS reran the previous models as well as several new models using the CHIS 2021-2022 ABS sample. Specifically, we appended all available data from voter registration databases, consumer databases, Marketing Systems Group (MSG) database information (namely, all ranges of surnames), and Census Planning Database (PDB) data, to the CHIS 2021 data. All these appended data serve as the independent variables (features) in random forest models, while the self-reported attributes (demographics, etc.) serve as the dependent variables. The CHIS 2022 data then served as the "test" data to evaluate the efficacy of the models on secondary data.

In addition to evaluating the predictive models, for CHIS 2023, we also investigated the utility of individual sample flags provided by MSG database information, including the surname flags, child indicator variables, and resident age information as well as PDB block-group characteristics including the density of households with African American residents and households with limited English proficiency.

Table 2-1 shows the attributes that we were trying to predict, whether we evaluated individual indicator flags, predictive models, or geo-demographic data, and which, if any, indicator was included in the final stratification.

Table 2-1: CHIS 2023 ABS Sample Stratification Predictors by Targeted Attribute

Targeted Attribute	Predictor Evaluated	Used in CHIS 2023
Vietnamese	Surname flag & predictive model	Surname flag
Korean	Surname flag & predictive model	Surname flag
Asian language Interview	Predictive model & geographic density indicator	Predictive model & geographic density indicator
Any Asian	Surname flag, predictive model, & geographic density indicator	Surname flag & predictive model
Spanish-language Interview	Predictive model & geographic density indicator	Predictive model & geographic density indicator
Hispanic	Surname flag, predictive model, & geographic density indicator	Surname flag
African American	Predictive models & geographic density indicator	Geographic density indicator
Native Hawaiian, Other Pacific Islander, American Indian, or Alaskan Native	Predictive model	Not specifically targeted
No H.S. Diploma	Predictive model	Not specifically targeted
Low income (under $35,000)	Predictive model	Not specifically targeted
Non-citizen	Predictive model	Not specifically targeted
Presence of children (under 19)	Sample frame flag & predictive model	Predictive model
Presence of adult aged 18-29	Sample frame flag & predictive model	Not specifically targeted
Presence of adult aged 65+	Sample frame flag & predictive model	Sample frame flag & predictive model

These identifying indicators were then used to create a hierarchical set of sample strata and each address in the selected sample was assigned to one of the strata with preference given to the higher ranked strata (for example, a household predicted to be Vietnamese was included in the Vietnamese strata regardless of which other categories they were predicted to be in). For CHIS 2023, the following strata were created:

1. Vietnamese
2. Korean
3. Likely Asian-language Interview
4. Likely Spanish-language interview
5. Hispanic
6. Other high-density non-English
7. Other Asian
8. High density African American
9. HH with children
10. Other 65+
11. Residual - Match
12. Residual – No match

This stratification scheme was designed to make use of the most effective predictive variables to target key demographic subgroups in an efficient way that minimizes the impact of the disproportionate sampling on the design effect. Those models that were not sufficiently predictive to add value were excluded[2]. It should be noted that this stratification includes two additional strata: 1) sample records for which none of the variables or models predicted any attribute, but for which auxiliary data could be matched to the address ("Residual - Match" sample) and sample for which no Big Data was found ("Residual - No match" sample). The final step in utilizing the models is to develop sampling fractions by which modeled households will be selected. The sample fractions below balanced the need to increase the frequency of the lowest incidence groups, while accounting for subgroups differences in response propensity and minimizing disproportionate weighting whenever possible. The relative sampling fractions were as follows:

1. Vietnamese (4.22)
2. Korean (2.88)
3. Likely Asian-language Interview (1.96)
4. Likely Spanish-language interview (2.95)

[2] Several targeted subgroups, including NHPI, AIAN, individuals with less than a high school diploma, individuals earning less than $35,000 per year, noncitizens, and younger adults did not yield models or predictor variables with enough predictive power to efficiently be included in the sample stratification scheme. However, as with CHIS 2021-2022 the current sample design aims for adequate representation of these subgroups.

5. Hispanic (1.46)
6. Other high-density non-English (1.50)
7. Other Asian (1.12)
8. High density African American (5.02)
9. HH with children (1.87)
10. Other 65+ (0.38)
11. Residual – Match (1.09)
12. Residual – No match (0.74)

Since the modelling is a post-generation process, we generated sample at the rate of the highest sampling fraction per geographic strata. Once the main sample was selected, random subsamples within modeled strata were drawn to achieve the final desired sampling fractions.

2.4 Oversampling

The CHIS design regularly includes additional samples for specialized analyses of certain geographic areas and hard-to-reach groups.

2.4.1 San Diego Oversample

As has been the case in prior years, San Diego County chose to oversample for additional statistical power in CHIS 2023. This geographic oversample targeted specific overall quotas by Health and Human Services Agency (HHSA) regions, for a total of 112 additional interviews in each region per year.

2.4.2 Prepaid Cell Oversample

Prepaid cell phone numbers are associated with cell phones that are "pay-as-you-go" and do not require a contract. Prepaid numbers are more likely to be used by Hispanics, people with lower education and lower income, and other related groups that are often underrepresented in general population samples (e.g., the uninsured). To better target populations not adequately covered under the ABS frame in CHIS 2023, we utilized a Prepaid cell oversample and targeted 450 prepaid oversample completes. This sample was targeted to reach in-language interviews, Hispanic and African American samples, and young adults aged 18 to 24. We used MSG listed cell sample and then further screened the sample to only retain interviews from in-language, Hispanic, African American, and 18-24 old respondents.

In 2023, SSRS released 113,812 sample pieces to meet the target for this oversample.

2.4.3 Long Beach Oversample

To provide researchers in City of Long Beach Health and Human Services with sufficient sample to produce estimates for a variety of health-related topics, in 2023 CHIS targeted an oversample of 500 interviews in the City of Long Beach from the following 11 Zip codes:

- 90802
- 90803
- 90804
- 90805
- 90806
- 90807
- 90808
- 90810
- 90813
- 90814
- 90815

We released 7,384 sample pieces to meet our target for this oversample.

2.4.4. Santa Clara Oversample

To provide Santa Clara County Public Health Department with sufficient samples to produce estimates for a variety of topics, CHIS 2023 sought to oversample 1,925 respondents from the county.

The Santa Clara oversample included an ABS stratified sample, with a similar stratification scheme as the production CHIS ABS sample.

Two key differences from the CHIS ABS sample were 1) the Korean stratum from the production CHIS ABS sample was replaced with a Filipino stratum, and 2) the sampling fractions were tweaked to help the county achieve desired subgroups sample sizes.

The strata and sampling fractions were as follows:

1. Vietnamese (1.27)
2. Filipino (4.42)
3. Likely Asian-language Interview (1.70)
4. Likely Spanish-language interview (3.92)
5. Hispanic (1.63)
6. Other high-density non-English (1.79)
7. Other Asian (0.88)

8. High density African American (2.74)
9. HH with children (0.77)
10. Other 65+ (0.37)
11. Residual – Match (1.42)
12. Residual – No match (0.75)

In addition, the geographic regions of East San Jose, Gilroy, Campbell, Cupertino, Milpitas, and Morgan Hill were oversampled, so that after being combined with the Santa Clara County data from production CHIS we would have a total of 400 completes in East San Jose and 100 completes each in the other specified regions.

A total of 26,118 sample pieces were released to meet the target for this oversample. As with the other oversamples, the sampled addresses were compared against the production CHIS sample and duplicates were purged. Due to the focused research analysis and the increased design effect in the targeted Santa Clara County the interviews completed from this oversample were not included in the publicly released CHIS data files.

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 2023 in Table 3-1.

Overall, CHIS 2023 was originally designed to yield 20,000 completed adult interviews. The targeted number of adolescent and child (proxy) interviews were established per projections from prior

CHIS waves. 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	20,000
Adolescents[1]	675
Children[1]	2,500
Supplemental samples (adults only):	
San Diego County	672
Long Beach	500
Prepaid Cell	450
Santa Clara County	1,925

Source: UCLA Center for Health Policy Research, 2023 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 2023 can be summarized as a stratified ABS design with strategic oversamples of households predicted to have certain attributes. As with previous waves, CHIS 2023 is a stratified design where study-eligible households were contacted, and one adult resident of the household (18 years of age or older) was chosen to be interviewed. Additionally, if the randomly chosen adult was the 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 previous recent waves, CHIS 2023 utilized 44 primary geographic strata as well as 8 Los Angeles-specific and 6 San Diego-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 2023 aimed to conduct a minimum of 100 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 2023 also sought to conduct a minimum of 20 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]
- 1.1 – Antelope Valley
- 1.2 – San Fernando Valley
- 1.3 – San Gabriel Valley
- 1.4 – Metro
- 1.5 – West
- 1.6 – South
- 1.7 – East
- 1.8 – South Bay

2 – San Diego (all)[2]
- 2.1 – North Coastal
- 2.2 – North Central
- 2.3 – Central
- 2.4 – South
- 2.5 – East
- 2.6 – North Inland

3 – Orange
4 – Santa Clara
5 – San Bernardino
6 – Riverside
7 – Alameda
8 – Sacramento
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
23 – San Luis Obispo
24 – Placer
25 – Merced
26 – Butte
27 – Shasta
28 – Yolo
29 – El Dorado
30 – Imperial
31 – Napa
32 – Kings
33 – Madera
34 – Monterey
35 – Humboldt
36 – Nevada
37 – Mendocino
38 – Sutter
39 – Yuba
40 – Lake
41 – San Benito
42 – Colusa, Glenn, Tehama
43 – Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity
44 – Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne

Source: UCLA Center for Health Policy Research, 2023 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, 2023 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, 2023 California Health Interview Survey.

3.3.1 Supplemental Samples for San Diego County

Similar to past cycles, San Diego County was oversampled to collect an additional 112 interviews in each of its six Health and Human Services Agency (HHSA) regions. This resulted in a target of 362 completes in each HHSA region, for a total target of 2,172 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 2023 adult interview targets including the San Diego County oversample are shown in Table 3-5. Table 3-5 also contains targets for prepaid cell oversample, the Long Beach oversample, and the Santa Clara County oversample.

Table 3-5. Initial and final 2023 targets for completed adult interviews by geographic strata (including supplemental samples)

Stratum	Initial Total [1,2]	Oversamples	Final Total [1,2]	Population size [3]
State Total	20,000		20,675	
1 Los Angeles (total)[1]	3,960		3,960	Over 9.8 million
1.1 – Antelope Valley	250		250	
1.2 – San Fernando Valley	835		835	
1.3 – San Gabriel Valley	686		686	
1.4 – Metro	439		439	
1.5 – West	250		250	
1.6 – South	394		394	
1.7 – East	508		508	
1.8 – South Bay	599		599	
2 San Diego (total)[2]	1,500	672	2,172	3.2 million or
2.1 – North Coastal	250	112	362	greater
2.2 – North Central	250	112	362	
2.3 – Central	250	112	362	
2.4 – South	250	112	362	
2.5 – East	250	112	362	
2.6 – North Inland	250	112	362	
3 Orange	1,230		1,230	900,000 to
4 Santa Clara	762		762	3.2 million
5 San Bernardino	764		764	
6 Riverside	866		866	
7 Alameda	685		685	
8 Sacramento	637		637	
9 Contra Costa	465		465	
10 Fresno	366		366	
11 San Francisco	437		437	600,000 to
12 Ventura	325		325	900,000
13 San Mateo	315		315	
14 Kern	321		321	
15 San Joaquin	271		271	
16 Sonoma	250		250	Medium
17 Stanislaus	250		250	counties
18 Santa Barbara	250		250	100,000 to
19 Solano	250		250	600,000
20 Tulare	250		250	
21 Santa Cruz	250		250	
22 Marin	250		250	

(continued)

Table 3-5. Initial and final 2023 targets for completed adult interviews by geographic strata (excluding supplemental samples) (continued)

Stratum	Initial Total [1,2b]	Oversamples	Final Total [1,2]	Population size [3]
23 San Luis Obispo	250		250	
24 Placer	250		250	
25 Merced	250		250	
26 Butte	250		250	
27 Shasta	250		250	
28 Yolo	250		250	
29 El Dorado	250		250	
30 Imperial	250		250	
31 Napa	250		250	
32 Kings	250		250	
33 Madera	250		250	
34 Monterey	250		250	
35 Humboldt	250		250	
36 Nevada	250		250	Small counties
37 Mendocino	250		250	Less than
38 Sutter	250		250	100,000
39 Yuba	250		250	
40 Lake	250		250	
41 San Benito	250		250	
42 Tehama, etc.	200		200	Small counties
43 Del Norte, etc.	200		200	combined
44 Amador, etc.	200		200	
Oversamples				
Prepaid Cell Oversample		450	450	
Long Beach Oversample		500	500	
Santa Clara County Oversample		1,925	1,925	

Source: UCLA Center for Health Policy Research, 2023 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 2023 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 by year, and it also enumerates sample with phone appends by modeled strata.

Table 3-6. Total ABS sample generated and fielded by modeled stratum

		2023 ABS Sample		
	Stratum	Modeled Households	Sample Mailed	Mailed Sample with Phone Appended
1.	Vietnamese	10,825	7,902	4,981
2.	Korean	18,689	10,963	6,664
3.	Likely Asian-language Interview	53,523	21,031	13,461
4.	Likely Spanish-language interview	183,369	114,431	75,282
5.	Hispanic	43,384	13,334	7,014
6.	Other high-density non-English	153,544	48,590	27,186
7.	Other Asian	29,442	5,681	3,854
8.	High density African American	14,306	13,373	8,342
9.	HH with children	136,550	51,780	35,469
10.	Other 65+	114,754	9,721	7,323
11.	Residual - Match	75,391	16,077	5,099
12.	Residual – No match	43,857	9,019	4,971

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.

The sample for CHIS 2023 was generated monthly and released in several waves throughout 2023. Each generation of sample in CHIS 2023 was de-duped with prior releases to ensure that addresses were not duplicated. Table 3-7 below shows the sample size and initial mailing date for each wave.

Table 3-7. ABS Sample releases by wave

	2023 ABS Sample	
Mail Wave	Sample Size	Initial Mailing
Main CHIS (Including San Diego Oversample)		
1	8,998	1/26/2023
2	14,500	3/2/2023
3	14,498	3/9/2023
4	14,500	3/16/2023
5	14,500	3/30/2023
6	14,497	3/30/2023
7	14,500	4/6/2023
8	14,499	4/13/2023
9	14,500	4/20/2023
10	14,497	4/27/2023
11	14,498	5/4/2023
12	14,500	5/11/2023
13	12,510	5/18/2023
14	12,510	5/25/2023
15	12,510	6/1/2023
16	12,509	6/8/2023
17	13,408	6/22/2023
18	13,408	6/29/2023
19	13,410	7/6/2023
20	13,410	7/13/2023
21	12,942	7/20/2023
22	17,158	8/3/2023
23	13,970	8/10/2023
24	5,669	8/24/2023
Long Beach Oversample		
51	3,304	4/27/2023
52	4,081	8/3/2023
Santa Clara Oversample		
55	13,152	5/11/2023
56	12,966	8/17/203

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.

4. WITHIN-HOUSEHOLD SAMPLING

In this section, 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 increase child (proxy) interviews in CHIS 2023 and describes the differential sampling across two child age groups. Section 4.3 provides a discussion of procedures for choosing an adolescent to interview.

4.1 Adult Sampling

As with previous waves of CHIS, adults are 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 from 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.
- 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

[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).

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

For CHIS, a child is defined 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 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 $n_2/(2n_1 + n_2)$. 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 2023.

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

Child selection probability	Age category of children in household[1]	CHIS 2023	
		n	pct
Equal	Only children 0 to 5 years	1,458	31%
	Only children 6 to 11 years	2,178	47%
Unequal	Children 0 to 5 and 6 to 11 years	995	21%
	Total	4,631	100%

Source: UCLA Center for Health Policy Research, 2023 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 discontinued this process since with the revised child-adult interview ordering, describe in section 4.2.2 every interview was in essence child first.

4.2.2 Child-then-Adult Ordering

To maximize the child sample size in CHIS 2019, the child rostering section was moved up to the end of Section A instead of Section G in the adult interview. 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. CHIS 2023 continued to follow this revised interview order of conducting the child interview before completing the adult interview.

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 at the end of 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 2023. 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 2023 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

Sample type/interview type	Completes	Targets	
	2023[1]	n	Pct[2]
Adult ABS	20,556	20,675	99.4%
Adult Listed Prepaid	169		
Child	3,203		
Child Listed Prepaid	22		
Teen	939		
Teen Listed Prepaid	3		
Prepaid Cell OS	452	450	100.4%
Long Beach OS	494	500	98.8%
Santa Clara County OS	2,026	1,925	105.2%

Source: UCLA Center for Health Policy Research, 2023 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 for 2023 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

Reported stratum	CHIS 2023				
	ABS Completes[1]	Prepaid Completes[1]	Total	Target	*% of target*
State-wide	20,556	169	20,725	20,675	100.2
1 Los Angeles	4,016	21	4,037	3,981	101.4
2 San Diego	2,175	11	2,186	2,172	100.6
3 Orange	1,205	0	1,205	1,244	96.9
4 Santa Clara	749	25	774	762	101.6
5 San Bernardino	761	0	761	758	100.4
6 Riverside	824	0	824	862	95.6
7 Alameda	659	1	660	681	96.9
8 Sacramento	573	20	593	626	94.7
9 Contra Costa	425	17	442	463	95.5
10 Fresno	364	0	364	362	100.6
11 San Francisco	427	0	427	434	98.4
12 Ventura	309	5	314	327	96.0
13 San Mateo	316	1	317	317	100.0
14 Kern	312	0	312	318	98.1
15 San Joaquin	249	6	255	268	95.1
16 Sonoma	222	15	237	250	94.8
17 Stanislaus	246	0	246	250	98.4
18 Santa Barbara	254	0	254	250	101.6
19 Solano	240	0	240	250	96.0
20 Tulare	252	0	252	250	100.8
21 Santa Cruz	232	9	241	250	96.4
22 Marin	259	0	259	250	103.6
23 San Luis Obispo	247	0	247	250	98.8
24 Placer	247	0	247	250	98.8
25 Merced	236	4	240	250	96.0

(continued)

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

Reported stratum	CHIS 2023				
	ABS Completes[1]	Prepaid Completes[1]	Total	Target	*% of target*
26 Butte	250	0	250	250	100.0
27 Shasta	261	1	262	250	104.8
28 Yolo	270	0	270	250	108.0
29 El Dorado	245	0	245	250	98.0
30 Imperial	269	0	269	250	107.6
31 Napa	266	0	266	250	106.4
32 Kings	240	0	240	250	96.0
33 Madera	259	0	259	250	103.6
34 Monterey	245	0	245	250	98.0
35 Humboldt	296	0	296	250	118.4
36 Nevada	266	0	266	250	106.4
37 Mendocino	224	12	236	250	94.4
38 Sutter	270	0	270	250	108.0
39 Yuba	238	0	238	250	95.2
40 Lake	262	0	262	250	104.8
41 San Benito	216	21	237	250	94.8
42 Tehama, etc.	236	0	236	200	118.0
43 Del Norte, etc.	204	0	204	200	102.0
44 Tuolumne, etc.	240	0	240	200	120.0

Source: UCLA Center for Health Policy Research, 2023 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. The reported stratum in these tables corresponds to the information provided by the screener respondent and is the same as reported in Table 5-2.

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

Reported Stratum	CHIS 2023		
	ABS Completes	Prepaid Completes	Total
State-wide	3,203	22	3,225
1 Los Angeles	631	2	633
2 San Diego	346	0	346
3 Orange	169	0	169
4 Santa Clara	113	3	116
5 San Bernardino	138	0	138
6 Riverside	115	0	115
7 Alameda	103	0	103
8 Sacramento	76	2	78
9 Contra Costa	72	4	76
10 Fresno	54	0	54
11 San Francisco	40	0	40
12 Ventura	41	1	42
13 San Mateo	55	0	55
14 Kern	68	0	68
15 San Joaquin	38	0	38
16 Sonoma	26	3	29
17 Stanislaus	44	0	44
18 Santa Barbara	44	0	44
19 Solano	26	0	26
20 Tulare	63	0	63
21 Santa Cruz	27	0	27
22 Marin	35	0	35
23 San Luis Obispo	28	0	28
24 Placer	42	0	42
25 Merced	56	0	56

(continued)

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

Reported Stratum	CHIS 2023		
	ABS Completes	Prepaid Completes	Total
26 Butte	45	0	45
27 Shasta	44	1	45
28 Yolo	33	0	33
29 El Dorado	31	0	31
30 Imperial	46	0	46
31 Napa	38	0	38
32 Kings	54	0	54
33 Madera	59	0	59
34 Monterey	49	0	49
35 Humboldt	39	0	39
36 Nevada	40	0	40
37 Mendocino	27	3	30
38 Sutter	37	0	37
39 Yuba	57	0	57
40 Lake	28	0	28
41 San Benito	34	3	37
42 Tehama, etc.	40	0	40
43 Del Norte, etc.	28	0	28
44 Tuolumne, etc.	24	0	24

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.

Table 5-6. Number of completed adolescent interviews by self-reported geographic stratum

Reported Stratum	CHIS 2023		
	ABS Completes	Prepaid Completes	Total
State-wide	939	3	942
1 Los Angeles	162	0	162
2 San Diego	85	0	85
3 Orange	38	0	38
4 Santa Clara	36	1	37
5 San Bernardino	54	0	54
6 Riverside	45	0	45
7 Alameda	25	0	25
8 Sacramento	25	1	26
9 Contra Costa	16	1	17
10 Fresno	19	0	19
11 San Francisco	10	0	10
12 Ventura	12	0	12
13 San Mateo	13	0	13
14 Kern	28	0	28
15 San Joaquin	12	0	12
16 Sonoma	14	0	14
17 Stanislaus	8	0	8
18 Santa Barbara	12	0	12
19 Solano	10	0	10
20 Tulare	16	0	16
21 Santa Cruz	12	0	12
22 Marin	15	0	15
23 San Luis Obispo	9	0	9
24 Placer	14	0	14
25 Merced	12	0	12

(Continued)

Table 5-7. Number of completed adolescent interviews by self-reported geographic stratum (continued)

	CHIS 2023		
Reported Stratum	ABS Completes	Prepaid Completes	Total
26 Butte	10	0	10
27 Shasta	17	0	17
28 Yolo	11	0	11
29 El Dorado	17	0	17
30 Imperial	20	0	20
31 Napa	12	0	12
32 Kings	18	0	18
33 Madera	18	0	18
34 Monterey	10	0	10
35 Humboldt	13	0	13
36 Nevada	13	0	13
37 Mendocino	12	0	12
38 Sutter	10	0	10
39 Yuba	10	0	10
40 Lake	7	0	7
41 San Benito	8	0	8
42 Tehama, etc.	10	0	10
43 Del Norte, etc.	9	0	9
44 Tuolumne, etc.	12	0	12

Source: UCLA Center for Health Policy Research, 2023 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. Starting with CHIS 2019, 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 assent (for a detailed discussion please refer to *Section 4.2* above).

Table 5-8. Number of completed interviews by interview combinations and sample

Interview combinations[1,2]	CHIS 2023 ABS		CHIS 2023 Listed Prepaid	
	n	pct	n	pct
Adult only	17,473	82.0	150	85.7
Adult and child	2,215	10.4	16	9.1
Adult and teen	601	2.8	3	1.7
Adult, child, and teen	267	1.3	0	0.0
Child only	691	3.2	6	3.4
Teen only	41	0.2	0	0.0
Child and teen only	30	0.1	0	0.0
Total	21,318	100.0	175	100.0

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.
Note: n = sample size; pct = unweighted percent.
[1] Includes completed and partial interviews
[2] Does not include counts for oversamples other than the San Diego oversample

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 modeled strata were very effective in targeting the groups of interest.

Table 6-1 shows the efficacy of the individual strata in reaching the target groups. Note that numbers in each row do not, necessarily, represent unique respondents 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). Similarly, the numbers in the columns do not add up to the amount in the total row, since a piece of sample could have been predicted to be in multiple groups. While sampled households were assigned to the modeled groups hierarchically (e.g., a piece of sample predicted to be a Vietnamese household and to have a household member aged 65+ would be assigned to the Vietnamese Household strata), for the purpose of evaluating the efficacy of the modeling on the final sample composition, in this section we present each modeled group independently. So, for instance, in tables 6-1 and 6-2, the piece of sample that is predicted to be Vietnamese and also to have a household member aged 65+, is presented in each of these respective predictor rows. The last row in tables 6-1 and 6-2 show the sample performance as a whole with respect to each targeted demographic.

The modeled strata performed best when targeting ethnic groups. For instance, of the adult completed interviews from the modeled Vietnamese group, 58.9% respondents self-identified as Vietnamese (Table 6-2). Of the adult completed interviews not in that group, 0.5% self-identified as Vietnamese. Similarly, the incidence of adult respondents who self-identified as Korean was 26.0% in the modeled Korean group and only 0.7% outside of that group.

In the “have children” modeled group, the incidence of the presence of a child or adolescent among adult respondents was 31.1%, which is an improvement over the incidence of the presence of a child or adolescent among adult respondents outside of that group (7.6%). Though the magnitude of improvement was the lowest among this modeled set, this modeling effort still proved effective in reaching the targeted group.

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

Model-Defined Group	Self-Reported Demographics							Completed Adult Interviews[1]
	Vietnamese	Korean	Other Asian	Hispanic / Spanish-Language Household	African American	Age 65+	Child / Adolescent Present in Household	
Vietnamese Household	322	25	156	8	1	167	117	547
Korean Household	65	298	634	25	11	353	264	1,147
Asian-Language Household	358	321	1,356	240	80	909	665	3,098
Spanish-Language Household	30	30	418	3,559	181	1,199	1,298	5,227
Hispanic Household	13	16	321	3,554	106	1198	1,176	4,725
Other non-English-Language Household	300	264	1,498	3,253	482	3,220	1,931	9,664
Other Asian Household	366	329	1,666	214	58	994	760	3,317
African American Household	31	30	274	606	452	816	439	2,379
Household with child (under 19)	287	254	1,625	3,790	545	1,383	3,580	11,496
Age 65+ Household	122	118	819	1,429	361	5,130	451	6,875
No auxiliary data group	20	46	263	446	105	420	336	1,599
Residual group	0	2	23	54	27	162	44	652
Completed Adult Interviews[1]	423	434	2,739	5,478	1,094	6,632	4,266	20,556

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.

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

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

Model-Defined Group	Self-Reported Demographics							Completed Adult Interviews[1]
	Vietnamese	Korean	Other Asian	Hispanic / Spanish-Language Household	African American	Age 65+	Child / Adolescent Present in Household	
Vietnamese Household	58.9%	4.6%	28.5%	1.5%	0.2%	30.5%	21.4%	547
Korean Household	5.7%	26.0%	55.3%	2.2%	1.0%	30.8%	23.0%	1,147
Asian-Language Household	11.6%	10.4%	43.8%	7.7%	2.6%	29.3%	21.5%	3,098
Spanish-Language Household	0.6%	0.6%	8.0%	68.1%	3.5%	22.9%	24.8%	5,227
Hispanic Household	0.3%	0.3%	6.8%	75.2%	2.2%	25.4%	24.9%	4,725
Other non-English-Language Household	3.1%	2.7%	15.5%	33.7%	5.0%	33.3%	20.0%	9,664
Other Asian Household	11.0%	9.9%	50.2%	6.5%	1.7%	30.0%	22.9%	3,317
African American Household	1.3%	1.3%	11.5%	25.5%	19.0%	34.3%	18.5%	2,379
Household with child (under 19)	2.5%	2.2%	14.1%	33.0%	4.7%	12.0%	31.1%	11,496
Age 65+ Household	1.8%	1.7%	11.9%	20.8%	5.3%	74.6%	6.6%	6,875
No auxiliary data group	1.3%	2.9%	16.4%	27.9%	6.6%	26.3%	21.0%	1,599
Residual group	0.0%	0.3%	3.5%	8.3%	4.1%	24.8%	6.7%	652
Completed Adult Interviews[1]	2.1%	2.1%	13.3%	26.6%	5.3%	32.3%	20.8%	20,556

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.

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

7. LIMITATIONS FOR CHIS SAMPLE DESIGN

There is the possibility of unmeasured error in this or any other population-based survey due to the survey's sample design and sample selection. The selected sample is one of many potential samples, and it is possible that the population parameters for each sample could vary slightly by random chance. While efforts were made to limit coverage error by including eligible cases and excluding ineligible cases, there is also the possibility that some eligible housing was absent from the frame.

8. 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 2023 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 2023 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, 2019-2020, 2021-2022, and 2023

County	2015-2016, 2017-2018, 2019-2020, 2021-2022, 2023 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, 2019-2020, 2021, and 2022 (continued)

County	2015-2016, 2017-2018, 2019-2020, 2023 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				
Lassen Modoc Plumas	43	43	43	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, 2023 California Health Interview Survey

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

Sampling stratum	CHIS 2023	
	Total	Sample with Appended phones
State-wide	321,902	199,646
1 Los Angeles	70,683	43,520
2 San Diego	31,203	19,987
3 Orange	17,298	10,979
4 Santa Clara	9,589	5,695
5 San Bernardino	15,428	9,965
6 Riverside	15,250	10,264
7 Alameda	7,393	4,564
8 Sacramento	7,504	4,681
9 Contra Costa	5,591	3,832
10 Fresno	6,014	3,696
11 San Francisco	5,244	2,962
12 Ventura	4,552	3,105
13 San Mateo	3,784	2,458
14 Kern	6,128	3,757
15 San Joaquin	5,293	3,480
16 Sonoma	2,334	1,423
17 Stanislaus	5,786	3,783
18 Santa Barbara	3,435	2,001
19 Solano	3,283	2,215
20 Tulare	5,628	3,417
21 Santa Cruz	2,449	1,525
22 Marin	2,555	1,696
23 San Luis Obispo	2,335	1,360
24 Placer	3,205	2,166
25 Merced	4,894	3,000
26 Butte	3,226	1,863
27 Shasta	3,318	2,011
28 Yolo	2,953	1,744
29 El Dorado	2,913	1,911

(continued)

Table A-3. Number of sample pieces selected by design stratum (continued)

Sampling stratum	CHIS 2023 Total	CHIS 2023 Sample with Appended phones
30 Imperial	5,620	3,245
31 Napa	3,530	2,200
32 Kings	5,786	3,566
33 Madera	6,462	3,758
34 Monterey	3,797	2,261
35 Humboldt	2,926	1,510
36 Nevada	2,952	1,731
37 Mendocino	3,009	1,517
38 Sutter	5,075	3,116
39 Yuba	4,901	3,025
40 Lake	4,231	2,156
41 San Benito	4,320	2,896
42 Tehama, etc.	4,142	2,373
43 Del Norte, etc.	2,746	1,391
44 Tuolumne, etc.	3,137	1,841

Source: UCLA Center for Health Policy Research, 2023 California Health Interview Survey.