

CHIS 2021-2022 Methodology Report Series

Report 1

Sample Design

September 2023



CALIFORNIA HEALTH INTERVIEW SURVEY

CHIS 2021-2022 METHODOLOGY SERIES

REPORT 1

SAMPLE DESIGN

SEPTEMBER 2023

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 2021-2022, 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 2021-2022 Methodology Series: Report 1 - Sample Design. Los Angeles, CA: UCLA Center for Health Policy Research, 2023.

Copyright © 2023 by the Regents of the University of California.

The California Health Interview Survey is a collaborative project of the UCLA Center for Health Policy Research with multiple funding sources. Funding for CHIS 2021-2022 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 Mental Health Services Authority, the California Tobacco Prevention Program, the California Wellness Foundation, First 5 California, Cedars-Sinai, The National Collaborative for Gun Violence Research, The National Bureau of Economic Research, Martin Luther King Community Healthcare, UCLA Asian-American Studies Center, and San Diego County Health and Human Services Agency.

PREFACE

Sample Design is the first in a series of methodological reports describing the 2021-2022 California Health Interview Survey (CHIS 2021-2022). 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 2021-2022 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 2021-2022

The methodological reports for CHIS 2021-2022 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 2021-2022. An appropriate sample design is a feature of a successful survey, and CHIS 2021-2022 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 2021-2022 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 <u>https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-methods/chis-methodology-reports-repository</u>. General information on CHIS data can be found on the California Health Interview Survey Web site at <u>http://www.chis.ucla.edu</u> or by contacting CHIS at <u>CHIS@ucla.edu</u>.

Table of Contents

1. CHIS 2021-2022 SAMPLE DESIGN AND METHODOLOGY SUMMARY	1-1
1.1 Overview	1-1
1.2 Sample Additions and Data Collection Methodology Updates	1-2
1.3 Sample Design Objectives	1-3
1.4 Data Collection	
1.5 Response Rates	
1.6 Weighting the Sample	
1.7 Imputation Methods	1-14
2. SAMPLING FRAMES AND METHODS	2-1
2.1. Addressed-Based Sampling Frame	2-1
2.2. Targeting Demographics through Predictive Modeling	2-1
2.3. Oversampling	
2.3.1. San Diego Oversample	
2.3.2 Cedars-Sinai Oversample	
2.3.3 Prepaid Cell Oversample	
2.3.4. American Indian and Alaskan Native (AIAN) Oversample	
2.3.5. Martin Luther King Community Healthcare (MLKCH) Oversample	
2.3.5. Santa Clara Oversample	
3. SAMPLING HOUSEHOLDS	
3.1 Population of Interest	
3.2 Analytic Objectives	
3.3 Sample Design	
3.3.1 Supplemental Samples for San Diego County	3-6
3.4 Sample Selection and Sample Releases	3-9
4. WITHIN-HOUSEHOLD SAMPLING	4-1
4.1 Adult Sampling	4-1
4.2 Child Sampling	
4.2.1 Child-First Procedure	
4.2.2 Child-then-Adult Ordering	
4.3 Adolescent Sampling	
5. ACHIEVED SAMPLE SIZES	
6. EVALUATING THE PREDICTIVE MODELING	6-1
7. LIMITATIONS FOR CHIS SAMPLE DESIGN	7-1
8. REFERENCES	8-1
APPENDIX A	9-1

List of Tables

Table 1-1. California county and county group strata used in the CHIS 2021-2022 sample design1	1-5
Table 1-2. Number of completed interviews by mode of interview and instrument ¹ 1	1-6
Table 1-3. CHIS 2021-2022 survey topic areas by instrument 1	1-8
Table 1-3. CHIS 2021-2022 survey topic areas by instrument (continued)1	1-9
Table 1-3. CHIS 2021-2022 survey topic areas by instrument (continued)1-	·10
Table 1-3. CHIS 2021-2022 survey topic areas by instrument (continued)1-	·11
Table 1-4a. CHIS response rates - Conditional1-	·12
Table 1-4b. CHIS response rates - Unconditional1-	·12
Table 2-1: CHIS 2021-2022 Sample Stratification Predictors by Targeted Attribute	2-2
Table 2-1: CHIS 2021-2022 Sample Stratification Predictors by Targeted Attribute (continued)2	2-3
Table 3-1. Initial targeted number of interviews by sample characteristics	3-2
Table 3-2. Geographic strata and sub-areas 3	3-4
Table 3-3. Los Angeles County Health Districts	
Table 3-4. Multi-county strata3	3-6
Table 3-5. Initial and final 2021-2022 targets for completed adult interviews by geographic strata (including supplemental samples)	3-7
Table 3-6. Initial and final 2021-2022 targets for completed adult interviews by geographic strata (excluding supplemental samples) (continued)	3-8
Table 3-6. Total sample generated and fielded by modeled stratum	3-9
Table 3-7. Sample release by wave	·10
Table 4-1. Distribution of households with children by child selection probability and year4	4-2
Table 5-1. Number of completed interviews by type of sample and year	5-1
Table 5-2. Number of completed adult interviews by self-reported stratum 5	5-2
Table 5-3. Number of completed child interviews by self-reported stratum	5-4
Table 5-4. Number of completed adolescent interviews by self-reported geographic stratum5	5-6
Table 5-5. Number of completed interviews by interview combinations and year	5-8
Table 6-1. Complete adult interviews by modeled variables (Counts)	5-2
Table 6-2. Competed adult interviews by modeled variables (percentages)	5-3
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	€-2
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)	€-3
Table A-2. Number of sample pieces selected by design stratum	€-4

1. CHIS 2021-2022 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 2021-2022.

- 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 <u>http://www.chis.ucla.edu</u> or contact CHIS at <u>CHIS@ucla.edu</u>. For methodology reports from previous CHIS cycles, go to <u>https://healthpolicy.ucla.edu/our-work/californiahealth-interview-survey-chis/chis-design-and-methods/chis-methodology-reports-repository.</u>

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 <u>https://healthpolicy.ucla.edu/our-work/publications</u> 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 2021-2022, 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.

The prepaid cell phone sample 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 2021-2022.

The CHIS design regularly includes additional samples for focused analysis of specific geographic areas or populations. The CHIS 2021-2022 included four oversamples:

- 1) In 2021 only, the Cedar-Sinai oversample was composed of ABS sample from LA County Service Planning Areas 1,2,4, and 5. These households were screened for Latinos or Asians who are aged 50 or older.
- 2) In both 2021 and 2022 American Indian and Alaska Natives (AIAN), were also oversampled in 2021. Respondents in this sample were asked in the screener whether they considered themselves to be American Indian or Alaska Native or to be of American Indian or Alaska Native decent.
- 3) CHIS 2022 oversampled households from 13 ZIP codes in LA County Service Planning Areas 6, 7, and 8 that surround the Martin Luther King Community Healthcare (MLKCH) hospital.

4) Lastly, CHIS 2022 oversampled Santa Clara County households.

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 <u>https://healthpolicy.ucla.edu/our-</u> work/training?keys=&gid%5B45%5D=45&sort_bef_combine=publish_date_DESC.

Additional documentation on constructing the CHIS sampling weights is available in the *CHIS* 2021-2022 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 2021-2022 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, as with CHIS 2019-2020, CHIS 2021-2022 continued to employ 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.

As with CHIS 2019-2020, the address-based sample in CHIS 2021-2022 was stratified into different strata that had higher incidences of individuals with targeted characteristics. For CHIS 2021-2022, 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 2021-2022 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 2021-2022, the following strata were created¹:

- 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 deigned 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 modelled households will be selected. The final sample fractions 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.

 $^{^{1}}$ The Santa Clara oversample employs a slightly different strata, please refer to Methodology Report 1 – Sample Design for additional details.

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. The child interview was completed by the randomly selected respondent who was the parent or guardian.

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

Table 1-1. California county and county group strata used in the CHIS 2021-2022 sample design

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

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 2021-2022, we utilized a Prepaid cell oversample and targeted 900 completes to obtain additional inlanguage interviews, Hispanic and African American samples, and young adults. 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 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 2021-2022, 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 2021-2022 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 2021-2022 by mode of interview. Note that these figures were accurate as of data collection completion for 2021-2022 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 <u>https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-methods/chis-design.</u>

	Adult	Child	Adolescent
Totals ²	46,810	7,505	2,177

Table 1-2. Number of completed interviews by mode of interview and instrument¹

Completes by Web	41,912	6,963	2,012
Completes by phone	4,898	542	165

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey. ¹ This table excludes the Santa Clara oversample.

² 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 72 minutes, 19 minutes, and 30 minutes for the adult, child, and adolescent interviews, respectively; the duration of the CAWI interviews averaged around 47 minutes, 13 minutes, and 21 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 83 minutes, 22 minutes, and 33 minutes for the adult, child, and adolescent interviews respectively; the non-English CAWI interviews had an average length of about 56 minutes, 16 minutes, and 23 minutes for the adult, child, and adolescent interviews were completed in a language other than English, as were about 13 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.

Health status	Adult	Adolescent	Child
General health status	\checkmark	\checkmark	✓
Days missed from work or school due to health problems	\checkmark	\checkmark	\checkmark
Health conditions	Adult	Adolescent	Child
Asthma	\checkmark	\checkmark	\checkmark
Diabetes, pre-diabetes/borderline diabetes	\checkmark		
Heart disease, high blood pressure	\checkmark		
Physical disability	\checkmark		
Mental health	Adult	Adolescent	Child
Mental health status	\checkmark	\checkmark	
Perceived need, access and utilization of mental health services	\checkmark	\checkmark	
Functional impairment, stigma	\checkmark		
Suicide ideation and attempts	\checkmark	\checkmark	
Mental health and technology	\checkmark	\checkmark	
Climate Change	\checkmark	\checkmark	
Health behaviors	Adult	Adolescent	Child
Dietary and nutritional intake, breastfeeding (younger than 3 years)	\checkmark		\checkmark
Sugar-sweetened beverages		\checkmark	\checkmark
Alcohol use, Cigarette use, E-cigarette use, Marijuana use, CBD use		\checkmark	
Opioid use	\checkmark		
Exposure to second-hand smoke/vapor, Exposure to marijuana	1		
smoke	•		
Sexual behaviors, HIV testing, HIV prevention medication	\checkmark	\checkmark	
Caregiving	\checkmark		
Gun Violence	Adult	Adolescent	Child
Firearm ownership/presence, loaded, and secure, firearm victimization, quick access to firearm	\checkmark	\checkmark	\checkmark
Women's health	Adult	Adolescent	Child
Pregnancy status/plans and birth control	\checkmark	\checkmark	
Intimate Partner violence	Adult	Adolescent	Child
Intimate partner violence	\checkmark		
Dental health	Adult	Adolescent	Child
Last dental visit, Main reason have not visited dentist, Number of dental visits, Location of dental service	\checkmark	\checkmark	√
Current dental insurance coverage	\checkmark		\checkmark
Condition of teeth	\checkmark	\checkmark	

(continued)

Neighborhood and housing	Adult	Adolescent	Child
Safety, social cohesion	✓	\checkmark	✓
Housing security/stability, length of residency	\checkmark		
Civic engagement, community involvement	\checkmark	\checkmark	
Encounters with police	\checkmark		
Adverse Childhood Experiences	Adult	Adolescent	Child
ACES Screener	\checkmark	\checkmark	
Past ACES screener	\checkmark	\checkmark	\checkmark
Positive Childhood Experiences	\checkmark	\checkmark	
Access to and use of health care	Adult	Adolescent	Child
Usual source of care, visits to medical doctor	\checkmark	\checkmark	\checkmark
Emergency room visits	\checkmark	\checkmark	\checkmark
Delays in getting care (prescriptions and medical care)	\checkmark	\checkmark	\checkmark
Communication problems with doctor	\checkmark		\checkmark
Contraception	\checkmark	\checkmark	
Timely appointment	\checkmark	\checkmark	\checkmark
Access to specialist and general doctors	\checkmark		
Tele-medical care	\checkmark		
Mammogram screening, colon cancer screening, HPV vaccination (only administered in Los Angeles Service Planning Areas 1, 2, 4, 5)	~		
Care coordination	\checkmark	\checkmark	\checkmark
Discrimination in healthcare setting	\checkmark		
Voter engagement	Adult	Adolescent	Child
Voter engagement	\checkmark		
Voter attitudes	\checkmark		
Food environment	Adult	Adolescent	Child
Availability of food in household over past 12 months, Hunger	\checkmark		
Health insurance	Adult	Adolescent	Child
Current insurance coverage, spouse's coverage, who pays for coverage	\checkmark	\checkmark	\checkmark
Health plan enrollment, characteristics and assessment of plan	\checkmark	\checkmark	\checkmark
Whether employer offers coverage, respondent/spouse eligibility	\checkmark		
Coverage over past 12 months, reasons for lack of insurance	\checkmark	\checkmark	\checkmark
High deductible health plans	\checkmark	\checkmark	\checkmark
Partial scope Medi-Cal, medical debt, hospitalizations	\checkmark		

Table 1-3. CHIS 2021-2022 survey topic areas by instrument (continued)

(continued)

Public program eligibility	Adult	Adolescent	Child
Household poverty level	✓		
Program participation (CalWORKs, Food Stamps, SSI, SSDI, WIC, TANF)	\checkmark	\checkmark	\checkmark
Assets, child support, Social security/pension, worker's compensation	\checkmark		
Medi-Cal eligibility, Medi-Cal renewal, Notice of actions from Medi-Cal	\checkmark		
Reason for Medi-Cal non-participation among potential beneficiaries	\checkmark	\checkmark	\checkmark
Use of public benefits among immigrant residents	\checkmark		
Parental involvement/adult supervision	Adult	Adolescent	Child
Parental involvement			\checkmark
Book ownership, source of reading materials, challenges to reading to child			✓
Child care and school	Adult	Adolescent	Child
Current child care arrangements			\checkmark
Paid child care	\checkmark		
First 5 California: Talk, Read, Sing Program / Kit for New Parents			\checkmark
Preschool/school attendance, school name		\checkmark	\checkmark
Preschool quality			\checkmark
Employment	Adult	Adolescent	Child
Employment status, spouse's employment status	✓		
Hours worked at all jobs	\checkmark		
Industry and occupation, firm size	\checkmark		
Paid Family Leave	\checkmark		
Income	Adult	Adolescent	Child
Respondent's and spouse's earnings last month before taxes	\checkmark		
Household income, number of persons supported by household income	\checkmark		

Table 1-3. CHIS 2021-2022 survey topic areas by instrument (continued)

(continued)

Respondent characteristics	Adult	Adolescent	Child
Race and ethnicity, age, gender, height, weight	✓	\checkmark	\checkmark
Veteran status	\checkmark		
Marital status, registered domestic partner status (same-sex couples)	\checkmark		
Sexual orientation	\checkmark		
Gender identity	\checkmark	\checkmark	
Gender expression		\checkmark	
Living with parents	\checkmark		
Education, English language proficiency	\checkmark		
Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home	\checkmark	\checkmark	\checkmark
COVID-19	Adult	Adolescent	Child
Ever though had COVID-19	\checkmark		
Ever tested positive for COVID-19	\checkmark		
COVID-19 vaccine status	\checkmark	\checkmark	\checkmark
Challenges experience due to COVID-19 pandemic	\checkmark		
Risk reduction practices	\checkmark		
Hate Incident (2022 only)	Adult	Adolescent	Child
Experienced hate incident	✓		

Table 1-3. CHIS 2021-2022 survey topic areas by instrument (continued)

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

1.5 Response Rates

The overall response rates for CHIS 2021-2022 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 2021-2022, the overall household response rate was 9.2 percent (the product of the screener response rate of 13.3 percent and the extended interview response rate at the household level of 69.5 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 2021-2022 Methodology Series: Report 4 – Response Rates*).

The extended interview response rate for the ABS sample varied across the adult (64.6 percent), child (82.5 percent) and adolescent (28.6 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 2021-2022 (see Table 1-4b).

Type of Sample	Screener ¹	Household (given screened) ¹	Adult (given screened) ¹	Child (given screened & eligibility) ¹	Adolescent (given screened & permission) ¹
Overall	13.3%	69.5%	64.6%	82.5%	28.6%

Table 1-4a. CHIS response rates - Conditional

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

¹The prepaid cell, Cedars-Sinai, MLKCH, Santa Clara, and AIAN oversamples are not included in these rates.

Type of Sample	Screener ¹	Household (given screened) ¹	Adult (given screened) ¹	Child (given screened & eligibility) ¹	Adolescent (given screened & permission) ¹
Overall	13.3%	9.2%	8.6%	10.9%	3.8%

Table 1-4b. CHIS response rates - Unconditional

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

¹ The prepaid cell, Cedars-Sinai, MLKCH, Santa Clara, and AIAN 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 "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 2021-2022, either a spouse/partner or adult child completed a proxy interview for twenty-two 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 <u>https://healthpolicy.ucla.edu/our-work/california-health-interview-survey-chis/chis-design-and-</u>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 2021-2022 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 2021-2022 were primarily created from the California Department of Finance's (DOF) 2021 and 2022 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 2021-2022 weights are based on 2010 Census counts, as were those used for the 2019-2020 cycle. Please pay close attention when comparing estimates using CHIS 2021-2022 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 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 2021-2022 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 2021-2022 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 nearcomplete 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 2021-2022 was selected via probability sampling methods and supplied by Marketing Systems Group (MSG).

2.2. 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.

2-1

For CHIS 2021-2022, we utilized the CHIS 2019-2020 data to test several models. 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 2019 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 2020 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 2021-2022, 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.

Targeted Attribute	Predictor Evaluated	Use in CHIS 2021-2022
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	Predictive model &	Predictive model &
Interview	geographic density indicator	geographic density indicator
Hispanic	Surname flag, predictive model, & geographic density indicator	Surname flag
African American	Predictive models & geographic density indicator	Geographic density indicator

Table 2-1: CHIS 2021-2022 Sample Stratification Predictors by Targeted Attribute

Targeted Attribute	Predictor Evaluated	Use in CHIS 2021-2022
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

Table 2-1: CHIS 2021-2022 Sample Stratification Predictors by Targeted Attribute (continued)

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 2021-2022, 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 excluded². 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 sampling fractions were as follows:

- 1. Vietnamese (3.67)
- 2. Korean (2.68)
- 3. Likely Asian-language Interview (2.0)
- 4. Likely Spanish-language interview (2.74)
- 5. Hispanic (1.04)
- 6. Other high-density non-English (1.43)
- 7. Other Asian (1.12)
- 8. High density African American (4.64)
- 9. HH with children (1.87)
- 10. Other 65 + (0.40)
- 11. Residual Match (0.98)
- 12. Residual No match (1.0)

Since the modelling is a post-generation process, we generated sample at the rate of the highest sampling fraction, which in this case is associated with the high-density African American stratum at a rate of 4.64. Once the main sample was selected, random subsamples within modeled strata were drawn to achieve the final desired sampling fractions. Specifically, under this design, all households included in the high-density African American stratum were selected for the study, but

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, the current sample design is estimated to yield an increased proportion of individuals from all these subgroups compared to CHIS 2019-2020.

only a relative sampling fraction of sample from other strata were selected relative to that in the African American stratum.

2.3. Oversampling

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

2.3.1. San Diego Oversample

As has been the case in prior years, San Diego County chose to oversample for additional statistical power in CHIS 2021-2022. The 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.

2.3.2 Cedars-Sinai Oversample

To provide researchers at Cedars-Sinai with sufficient sample to produce estimates for a variety of cancer screening questions, CHIS 2021 oversampled 800 Latinos and Asians aged 50 and older in LA County SPAs 1,2,4, and 5.

The Cedars-Sinai oversample was an ABS stratified sample, where relevant surname flags were used to model three strata, an Asian stratum, a Hispanic stratum, and a residual stratum. We oversample the relevant strata with a maximum random sampling fraction of 5. The sampled addresses were compared against the main CHIS sample and duplicates were purged. The sample was further screened, so that only interviews from Latinos and Asians aged 50 and older in LA County SPAs 1,2,4, and 5 were retained.

We released 91,436 sample pieces to meet our target for this oversample. Due to the focused research analysis and the increased design effect in the targeted Los Angeles County SPAs, the interviews completed from this oversample were not included in the publicly released CHIS data files.

2.3.3 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

2-5

the ABS frame in CHIS 2021-2022, we utilized a Prepaid cell oversample and targeted 900 prepaid oversample completes, 450 completes in each year. 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 inlanguage, Hispanic, African American, and 18-24 old respondents.

Across the two years, we released 208,628 sample pieces to meet our target for this oversample.

2.3.4. American Indian and Alaskan Native (AIAN) Oversample

CHIS 2021-2022 sought to conduct an additional oversample of 250 adults, 125 in each 2021 and 2022, who identify as American Indian or Alaska Native and live in rural areas. The sample for this oversample was an address-based sample. Using a rural definition of at least 75% of the Block Group addresses being classified as rural, the sample design focused on Block Groups that were a minimum AIAN density and also meet the 75% rural definition. For CHIS 2021 the minimum threshold for a block group to be oversampled was to have a 4.2% AIAN density. This threshold was lowered to 3.2% in July 2022. The sampled addresses were compared against the main CHIS sample and duplicates were purged.

We released 50,428 sample pieces to meet our target for this oversample.

2.3.5. Martin Luther King Community Healthcare (MLKCH) Oversample

CHIS 2022 also oversampled 400 adults from 13 ZIP codes in LA county SPAs 6, 7, 8 that surround MLKCH hospital. This oversample used the same model-based stratification as the main ABS sample, while sampling proportionally from the 13 ZIP codes.

We released 15,771 sample pieces to meet our target for this oversample.

2.3.5. Santa Clara Oversample

To provide Santa Clara County with sufficient samples to produce estimates for a variety of topics, CHIS 2022 oversampled 1,925 respondents from the county.

Like production CHIS, the Santa Clara oversample was an ABS stratified sample, with a similar stratification scheme as the production CHIS ABS sample. The two key differences 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 (0.50)
- 2. Filipino (4.93)
- 3. Likely Asian-language Interview (0.59)
- 4. Likely Spanish-language interview (2.34)
- 5. Hispanic (0.44)
- 6. Other high-density non-English (0.76)
- 7. Other Asian (0.81)
- 8. High density African American (0.52)
- 9. HH with children (0.60)
- 10. Other 65+ (0.90)
- 11. Residual Match (0.65)
- 12. Residual No match (0.62)

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.

We released 27,750 sample pieces to meet our 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 2021-2022 in Table 3-1.

Overall, CHIS 2021-2022 was originally designed to yield 40,000 completed adult interviews. The targeted number of adolescent and child (proxy) interviews were established per projections from

3-1

CHIS 2019-2020. Targets by design strata and for the supplemental samples are discussed in detail in Section 3.3.

Characteristics	Interviews (n)	
State-wide, Main Study, Overall		
Adults (overall)	40,000	
Adolescents (overall) ¹	1,350	
Children (overall) ¹	4,900	
Supplemental samples (adults only):		
San Diego County (2021-2022)	1,344	
Cedars-Sinai (2021)	800	
Prepaid Cell (2021-2022)	900	
AIAN (2021-2022)	250	
MLKCH (2022)	400	
Santa Clara County (2022)	1,925	

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

Source: UCLA Center for Health Policy Research, 2021-2022 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 2021-2022 can be summarized as a stratified ABS design with strategic oversamples of households predicted to have certain attributes. As with previous waves, CHIS 2021-2022 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 2021-2022 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 2021-2022 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 2021-2022 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 $(all)^1$	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,
8 – Sacramento	Siskiyou, Trinity
9 – Contra Costa	44 - Amador, Alpine, Calaveras, Inyo, Mariposa,
10 – Fresno	Mono, Tuolumne
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	
urce: UCLA Center for Health Policy Research, 20	021-2022 California Health Interview

Source: UCLA Center for Health Policy Research, 2021-2022 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.

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: LICL & Center for Health Policy Research, 2021-2022 California Health Interv	iew Su

 Table 3-3. Los Angeles County Health Districts

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

Table 3-4. Multi-county strata

42 - Tehama,	etc.
42.1 - Tel	hama
42.2 - Gle	enn
42.3 - Co	lusa
43 - Del Nort	e, etc.
43.1 - De	l Norte
43.2 - Sis	skiyou
43.3 - Las	ssen
43.4 - Tri	nity
43.5 - Mo	odoc
43.6 - Plu	imas
43.7 - Sie	erra
44 - Tuolumn	ie, etc.
44.1 - Tu	olumne
44.2 - Ca	laveras
44.3 - An	nador
44.4 - Iny	/0
44.5 - Ma	ariposa
44.6 - Mo	- DNO
44.7 - Al	pine
Source: UCLA	Center for Health Policy Research, 2021-2022 California Health Interview

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

3.3.1 Supplemental Samples for San Diego County

As with the CHIS 2019-2020, 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 2021-2022 adult interview targets including the San Diego County oversample are shown in Table 3-5. Table 3-5 also contains targets for the Cedars-Sinai Oversample, prepaid cell oversample, the AIAN oversample, the MLKCH oversample and the Santa Clara County oversample.

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	1,344	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	1,524		1,524	3.2 million
5 San Bernardino	1,528		1,528	
6 Riverside	1,732		1,732	
7 Alameda	1,370		1,370	
8 Sacramento	1,274		1,274	
9 Contra Costa	930		930	
10 Fresno	732		732	
11 San Francisco	874		874	600,000 to
12 Ventura	650		650	900,000
13 San Mateo	630		630	
14 Kern	642		642	
15 San Joaquin	542		542	
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	

Table 3-5. Initial and final 2021-2022 targets for completed adult interviews by geographic strata (including supplemental samples)

(continued)

Stratum	Initial Total	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 Amador, etc.	400		400	
Cedars-Sinai OS		800	800	
Prepaid Cell OS		900	900	
AIAN OS		250	250	
MLKCH OS		400	400	
Santa Clara County OS		1,92	5 1,925	5

Table 3-6. Initial and final 2021-2022 targets for completed adult interviews by geographic strata (excluding supplemental samples) (continued)

Source: UCLA Center for Health Policy Research, 2021-2022 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.

³Based 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 by year, and it also enumerates sample with phone appends by modeled strata.

		2021			2022	
Stratum	Modeled Households	Sample Mailed	Mailed Sample with Phone Appended	Modeled Households	Sample Mailed	Mailed Sample with Phone Appended
1. Vietnamese	12,148	8,306	5,922	9,733	7,477	4,050
2. Korean	20,314	10,259	6,930	15,863	9,889	5,156
3. Likely Asian- language Interview	60,231	22,501	16,206	48,634	21,026	11,308
4. Likely Spanish- language interview	198,188	102,786	74,529	165,412	105,776	57,533
5. Hispanic	52,589	10,483	6,336	34,515	11,764	4,924
6. Other high-density non-English	251,555	67,187	44,013	193,092	68,103	30,780
7. Other Asian	17,380	3,631	2,798	12,914	3,051	1,792
8. High density African American	21,517	18,938	13,822	20,284	18,859	9,683
9. HH with children	132,443	45,127	34,439	90,222	39,904	22,624
10. Other 65+	124,458	9,283	8,092	82,058	8,486	5,883
11. Residual - Match	64,727	12,124	5,468	40,045	10,377	2,397
12. Residual – No match	48,824	8,706	5,985	33,190	8,627	4,037

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

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

The sample for CHIS 2021-2022 was generated monthly and released in several waves as noted in table 3-6. Each generation of sample in CHIS 2021-2022 was de-duped with prior releases to ensure that addresses were not duplicated.

Table 3-7 below shows the sample size and mailing date for each wave.

Table 3-7. Sample release by wa	ve
---------------------------------	----

	2	2021		022
Mail Wave	Sample Size	Initial Mailing	Sample Size	Initial Mailing
1	13,860	3/15/2021	12,266	2/10/2022
2	13,858	3/15/2021	12,268	2/10/2022
3	13,862	3/18/2021	12,268	2/17/2022
4	13,859	3/25/2021	12,268	2/24/2022
5	13,862	4/1/2021	12,269	3/3/2022
6	13,861	4/8/2021	11,887	3/10/2022
7	13,862	4/15/2021	11,887	3/17/2022
8	13,859	4/22/2021	11,887	3/24/2022
9	13,862	4/29/2021	11,886	3/31/2022
10	13,861	5/6/2021	11,886	4/7/2022
11	13,861	5/13/2021	11,887	4/14/2022
12	13,863	5/20/2021	11,886	4/21/2022
13	13,862	5/27/2021	11,889	4/28/2022
14	13,860	6/3/2021	11,060	5/12/2022
15	13,862	6/10/2021	11,059	5/19/2022
16	13,861	6/17/2021	14,733	6/2/2022
17	13,111	6/24/2021	13,278	6/16/2022
18	12,606	7/1/2021	13,277	6/16/2022
19	12,281	7/8/2021	13,277	6/23/2022
20	11,973	7/15/2021	12,931	6/30/2022
21	12,491	7/22/2021	16,194	7/14/2022
22	12,490	7/29/2021	14,490	7/21/2022
23	7,744	8/12/2021	13,984	7/28/2022
24	4,944	8/19/2021	12,856	8/4/2022
25	9,916	9/2/2021	9,766	8/18/2022

Source: UCLA Center for Health Policy Research, 2021-2022 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 boost child (proxy) interviews in CHIS 2021-2022 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 survey³. 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 2021-2022.

Child selection probability	Age category of children in household ¹	CHIS 2021		CHIS 2022		Total	
		n	pct	n	pct	n	pct
Equal	Only children 0 to 5 years	1628	31%	1,270	31%	2,898	31%
	Only children 6 to 11 years	2,443	47%	1,894	47%	4,337	47%
Unequal	Children 0 to 5 and 6 to 11 years	1,102	21%	908	22%	2,010	22%
	Total	5,173	100%	4,072	100%	9,245	100%

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

Source: UCLA Center for Health Policy Research, 2021-2022 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 2021-2022 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 2021-2022 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 2021-2022. 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 2021-2022 Methodology Series: Report 4 – Response Rates*.

Table 5-1 compares the number of completed interviews by interview type.

Sample type/interview	Completes	Tar	gets	Completes	Та	rgets	Completes	s Ta	argets
type	2021	n	pct ^b	2022	n	pct ^b	Total	n	pct ^b
Adult	23,8161	20,676	115.2%	20,3131	20,676	98.2%	44,129 ¹	41,352	106.7%
Child	3,931			3,156			7,087		
Teen	1,159			942			2,101		
Cedars-Sinai OS	8941	800	111.8%				894 ¹	800	111.8%
Prepaid Cell OS	503 ¹	500	100.6%	543 ¹	450	120.7%	1,0461	950	110.1%
AIAN OS	1341	125	107.2%	1211	125	96.8%	255 ¹	250	102.0%
MLKCH OS				4861	400	121.5%	4861	400	121.5%
Santa Clara County OS				2,0611	1925	107.1%	2,0611	1,925	107.1%

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

Source: UCLA Center for Health Policy Research, 2021-2022 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 2021 divided by the target within sample type.

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

	C	HIS 2021		(CHIS 2022			Total	
Reported stratum	Actual Completes	Target	% of target	Actual Completes	Target	% of target	Actual Completes	Target	% of target
State-wide	23,816	20,675	115.2	20,313	20,675	98.2	44,129	41,350	106.7
1 Los Angeles	4,696	3,981	118.0	3,883	3,981	97.5	8,579	7,962	107.7
2 San Diego	2,523	2,172	116.2	2,199	2,172	101.2	4,722	4,344	108.7
3 Orange	1,374	1,244	110.5	1,185	1,244	95.3	2,559	2,488	102.9
4 Santa Clara	921	762	120.9	740	762	97.1	1,661	1,524	109.0
5 San Bernardino	850	758	112.1	746	758	98.4	1,596	1,516	105.3
6 Riverside	956	862	110.9	809	862	93.9	1,765	1,724	102.4
7 Alameda	898	681	131.9	652	681	95.7	1,550	1,362	113.8
8 Sacramento	729	626	116.5	611	626	97.6	1,340	1,252	107.0
9 Contra Costa	533	463	115.1	450	463	97.2	983	926	106.2
10 Fresno	402	362	111.0	371	362	102.5	773	724	106.8
11 San Francisco	558	434	128.6	430	434	99.1	988	868	113.8
12 Ventura	342	327	104.6	332	327	101.5	674	654	103.1
13 San Mateo	370	317	116.7	300	317	94.6	670	634	105.7
14 Kern	335	318	105.3	324	318	101.9	659	636	103.6
15 San Joaquin	291	268	108.6	275	268	102.6	566	536	105.6
16 Sonoma	266	250	106.4	244	250	97.6	510	500	102.0
17 Stanislaus	261	250	104.4	250	250	100.0	511	500	102.2
18 Santa Barbara	293	250	117.2	247	250	98.8	540	500	108.0
19 Solano	291	250	116.4	251	250	100.4	542	500	108.4
20 Tulare	283	250	113.2	251	250	100.4	534	500	106.8
21 Santa Cruz	275	250	110.0	256	250	102.4	531	500	106.2
22 Marin	321	250	128.4	252	250	100.8	573	500	114.6

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

(Continued)

	CH	IS 2021		C	CHIS 2022			Total	
Reported stratum	Actual Completes	Target	% of target	Actual Completes	Target	% of target	Actual Completes	Target	% of target
23 San Luis Obispo	307	250	122.8	232	250	92.8	539	500	107.8
24 Placer	275	250	110.0	253	250	101.2	528	500	105.6
25 Merced	283	250	113.2	239	250	95.6	522	500	104.4
26 Butte	280	250	112.0	221	250	88.4	501	500	100.2
27 Shasta	275	250	110.0	219	250	87.6	494	500	98.8
28 Yolo	341	250	136.4	228	250	91.2	569	500	113.8
29 El Dorado	272	250	108.8	241	250	96.4	513	500	102.6
30 Imperial	315	250	126.0	231	250	92.4	546	500	109.2
31 Napa	269	250	107.6	237	250	94.8	506	500	101.2
32 Kings	310	250	124.0	237	250	94.8	547	500	109.4
33 Madera	287	250	114.8	263	250	105.2	550	500	110.0
34 Monterey	259	250	103.6	257	250	102.8	516	500	103.2
35 Humboldt	295	250	118.0	236	250	94.4	531	500	106.2
36 Nevada	230	250	92.0	271	250	108.4	501	500	100.2
37 Mendocino	291	250	116.4	235	250	94.0	526	500	105.2
38 Sutter	252	250	100.8	272	250	108.8	524	500	104.8
39 Yuba	265	250	106.0	261	250	104.4	526	500	105.2
40 Lake	271	250	108.4	258	250	103.2	529	500	105.8
41 San Benito	320	250	128.0	250	250	100.0	570	500	114.0
42 Tehama, etc.	208	200	104.0	224	200	112.0	432	400	108.0
43 Del Norte, etc.	232	200	116.0	191	200	95.5	423	400	105.8
44 Tuolumne, etc.	211	200	105.5	199	200	99.5	410	400	102.5

Source: UCLA Center for Health Policy Research, 2021-2022 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.

	CHIS 2021	CHIS 2022	Total
State-wide	3,931	3,156	7,087
1 Los Angeles	779	594	1,373
2 San Diego	427	322	749
3 Orange	213	186	399
4 Santa Clara	154	134	288
5 San Bernardino	167	145	312
6 Riverside	184	119	303
7 Alameda	158	122	280
8 Sacramento	105	89	194
9 Contra Costa	75	55	130
10 Fresno	94	60	154
11 San Francisco	64	42	106
12 Ventura	56	50	106
13 San Mateo	68	52	120
14 Kern	57	74	131
15 San Joaquin	53	36	89
16 Sonoma	32	26	58
17 Stanislaus	52	36	88
18 Santa Barbara	38	33	71
19 Solano	31	30	61
20 Tulare	62	60	122
21 Santa Cruz	41	33	74
22 Marin	52	39	91
23 San Luis Obispo	36	32	68
24 Placer	56	46	102
25 Merced	48	51	99
26 Butte	61	29	90
27 Shasta	39	27	66
28 Yolo	58	33	91
29 El Dorado	40	33	73
30 Imperial	76	39	115
31 Napa	28	39	67
32 Kings	67	72	139
33 Madera	54	44	98

(Continued)

	CHIS 2019	CHIS 2022	Total
34 Monterey	45	44	89
35 Humboldt	36	31	67
36 Nevada	31	29	60
37 Mendocino	31	29	60
38 Sutter	36	39	75
39 Yuba	55	45	100
40 Lake	30	29	59
41 San Benito	52	48	100
42 Tehama, etc.	37	35	72
43 Del Norte, etc.	30	22	52
44 Tuolumne, etc.	23	23	46

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

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

	CHIS 2021	CHIS 2022	Total
State-wide	1,159	942	2,101
1 Los Angeles	212	170	382
2 San Diego	117	107	224
3 Orange	68	50	118
4 Santa Clara	39	27	66
5 San Bernardino	44	49	93
6 Riverside	51	35	86
7 Alameda	44	34	78
8 Sacramento	31	20	51
9 Contra Costa	26	21	47
10 Fresno	27	19	46
11 San Francisco	13	9	22
12 Ventura	16	13	29
13 San Mateo	16	13	29
14 Kern	25	27	52
15 San Joaquin	12	11	23
16 Sonoma	8	10	18
17 Stanislaus	17	13	30
18 Santa Barbara	17	13	30
19 Solano	11	5	16
20 Tulare	14	11	25
21 Santa Cruz	14	9	23
22 Marin	18	13	31
23 San Luis Obispo	12	6	18
24 Placer	16	16	32
25 Merced	19	16	35
26 Butte	14	13	27
27 Shasta	17	7	24
28 Yolo	25	7	32
29 El Dorado	13	16	29
30 Imperial	27	16	43
31 Napa	10	8	18
32 Kings	19	18	37
33 Madera	14	19	33

(Continued)

	CHIS 2021	CHIS 2022	Total
34 Monterey	11	9	20
35 Humboldt	17	8	25
36 Nevada	12	15	27
37 Mendocino	6	7	13
38 Sutter	11	15	26
39 Yuba	21	11	32
40 Lake	11	9	20
41 San Benito	17	13	30
42 Tehama, etc.	9	14	23
43 Del Norte, etc.	11	7	18
44 Tuolumne, etc.	7	13	20

Source: UCLA Center for Health Policy Research, 2021-2022 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).

Interview combinations ^{1,2}	CHIS 2021		CHIS 2022		Total	
inconview combinations	n	pct	n	pct	n	pct
Adult only	20,035	81.0	17,408	82.2	37,443	81.5
Adult and child	2,694	10.9	801	3.8	3,495	7.6
Adult and teen	723	2.9	49	0.2	772	1.7
Adult, child, and teen	364	1.5	2,034	9.6	2,398	5.2
Child only	844	3.4	572	2.7	1,416	3.1
Teen only	43	0.2	22	0.1	65	0.1
Child and teen only	29	0.1	299	1.4	328	0.7
Total	24,732	100.0	21,185	100.0	45,917	100.0

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

Source: UCLA Center for Health Policy Research, 2021-2022 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, 731 respondents self-identified as Vietnamese (Table 6-1). Of the adult completed interviews not in that group, 224 respondents self-identified as Vietnamese. Additionally, the total number of adult completed interviews from the modeled Vietnamese stratum was 1,306, while the total number of adult completed interviews not in that stratum was 42,823, so the incidence of adult respondents who self-identified as Vietnamese was 56.0% in the modeled Vietnamese group and only 0.5% outside of that group (Table 6-2). Similarly, the incidence of adult respondents who self-identified as Korean was 23.4% in the modeled Korean group and only 0.6% outside of that group.

In the "have children" modeled group, the incidence of the presence of a child or adolescent among adult respondents was 29.8%, which is an improvement over the incidence of the presence of a child or adolescent among adult respondents outside of that group (10.1%). Though the magnitude of improvement was the lowest among this modeled set, this modeling effort still proved effective in reaching the targeted group.

6-1

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	731	49	392	52	7	298	308	1,306
Korean Household	154	609	1,419	89	32	690	596	2,600
Asian- Language Household	786	602	2,896	615	159	1,815	1,449	6,691
Spanish- Language Household	55	43	779	7,010	292	2,077	2,616	10,020
Hispanic Household	28	28	605	7,206	179	2,022	2,544	9,356
Other non- English- Language Household	620	548	3,411	6,927	1,100	6,561	4,268	20,745
Other Asian Household	805	658	3,461	1,012	187	1,899	1,725	7,552
African American Household	79	65	574	1,350	1,066	1,846	998	5,421
Household with child (under 19)	659	496	3,576	7,690	968	2,772	7,280	24,392
Age 65+ Household	249	211	1,570	2,660	760	10,327	801	13,698
No auxiliary data group	5	10	91	161	46	380	181	1,124
Residual group	4	6	61	181	55	256	161	1,431
Completed Adult Interviews ¹	955	872	5,982	11,454	2,316	13,408	9,278	44,129

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

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.

¹ Includes interviews meeting the criteria as partially complete.

Model- Defined Group		Self-Reported Demographics							
	Vietnamese	Korean	Other Asian	Hispanic / Spanish- Language Household	African American	Age 65+	Child / Adolescent Present in Household		
Vietnamese Household	56.0%	3.8%	30.0%	4.0%	0.5%	22.8%	23.6%	1,306	
Korean Household	5.9%	23.4%	54.6%	3.4%	1.2%	26.5%	22.9%	2,600	
Asian- Language Household	11.7%	9.0%	43.3%	9.2%	2.4%	27.1%	21.7%	6,691	
Spanish- Language Household	0.5%	0.4%	7.8%	70.0%	2.9%	20.7%	26.1%	10,020	
Hispanic Household	0.3%	0.3%	6.5%	77.0%	1.9%	21.6%	27.2%	9,356	
Other non- English- Language Household	3.0%	2.6%	16.4%	33.4%	5.3%	31.6%	20.6%	20,745	
Other Asian Household	10.7%	8.7%	45.8%	13.4%	2.5%	25.1%	22.8%	7,552	
African American Household	1.5%	1.2%	10.6%	24.9%	19.7%	34.1%	18.4%	5,421	
Household with child (under 19)	2.7%	2.0%	14.7%	31.5%	4.0%	11.4%	29.8%	24,392	
Age 65+ Household	1.8%	1.5%	11.5%	19.4%	5.5%	75.4%	5.8%	13,698	
No auxiliary data group	0.4%	0.9%	8.1%	14.3%	4.1%	33.8%	16.1%	1,124	
Residual group	0.3%	0.4%	4.3%	12.6%	3.8%	17.9%	11.3%	1,431	
Completed Adult Interviews ¹	2.2%	2.0%	13.6%	26.0%	5.2%	30.4%	21.0%	44,129	

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

Source: UCLA Center for Health Policy Research, 2021-2022 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. <u>https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx</u>.
- 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). <u>http://www.surveypractice.org/index.php/SurveyPractice/article/view/264/html_8</u>
- 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 addressbased 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 2021-2022 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 2021-2022 separately by design stratum.

County	2015-2016, 2017-2018, 2019- 2020, 2021-2022 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	34

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, 2021-2022 Strata	2013-2014 Strata	2005, 2007, 2009, 2011-2012 Strata	2001, 2003 Strata
Lake	40	40	40	37
Mendocino	37	37	37	57
Sutter	38	38	38	39
Yuba	39	39	39	39
Colusa				
Glenn	42	42	42	38
Tehama				
Nevada	36	36	36	40
Humboldt	35	35	35	25
Del Norte				35
Lassen				
Modoc		43		36
Plumas	43	43	43	
Sierra				40
Trinity				40
Siskiyou		43.2		36
Amador				
Alpine				
Inyo		44		
Mariposa	44		44	41
Mono				
Tuolumne		44.1		
Calaveras	-	44.2	_	

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)

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey

	CHI	S 2021	CHI	S 2022	Т	otal
Sampling stratum	Total	Sample with Appended phones	Total	Sample with Appended phones	Total	Sample with Appended phones
State-wide	319,331	224,540	313,339	160,167	632,670	384,707
1 Los Angeles	70,306	49,407	69,772	35,868	140,078	85,275
2 San Diego	31,259	22,364	30,641	16,081	61,900	38,445
3 Orange	16,602	12,120	16,439	8,829	33,041	20,949
4 Santa Clara	9,732	6,639	8,647	4,377	18,379	11,016
5 San Bernardino	13,806	10,162	15,468	8,057	29,274	18,219
6 Riverside	14,864	11,304	15,339	8,266	30,203	19,570
7 Alameda	8,624	6,153	7,309	3,782	15,933	9,935
8 Sacramento	8,041	5,819	7,627	3,990	15,668	9,809
9 Contra Costa	5,958	4,549	5,277	3,051	11,235	7,600
10 Fresno	7,095	4,873	7,242	3,755	14,337	8,628
11 San Francisco	5,468	3,628	4,797	2,282	10,265	5,910
12 Ventura	4,119	3,205	4,480	2,530	8,599	5,735
13 San Mateo	3,796	2,729	3,750	2,053	7,546	4,782
14 Kern	6,029	4,204	6,394	3,148	12,423	7,352
15 San Joaquin	4,820	3,531	5,546	2,961	10,366	6,492
16 Sonoma	2,813	1,960	2,723	1,350	5,536	3,310
17 Stanislaus	4,486	3,221	4,806	2,600	9,292	5,821
18 Santa Barbara	3,100	2,152	3,233	1,536	6,333	3,688
19 Solano	3,637	2,823	3,247	1,874	6,884	4,697
20 Tulare	4,878	3,396	4,959	2,185	9,837	5,581
21 Santa Cruz	2,801	1,913	2,890	1,411	5,691	3,324
22 Marin	2,856	2,095	2,554	1,396	5,410	3,491
23 San Luis Obispo	2,615	1,771	2,295	1,089	4,910	2,860
24 Placer	2,935	2,268	3,036	1,721	5,971	3,989
25 Merced	5,247	3,502	4,936	2,345	10,183	5,847
26 Butte	3,140	2,070	2,726	1,274	5,866	3,344
27 Shasta	3,378	2,319	2,532	1,293	5,910	3,612
28 Yolo	2,815	1,853	2,587	1,253	5,402	3,106
29 El Dorado	2,823	2,085	2,906	1,493	5,729	3,578

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

(Continued)

	CH	IS 2021	CH	IS 2022]	Total
Sampling stratum	Total	Sample with Appended phones	Total	Sample with Appended phones	Total	Sample with Appended phones
30 Imperial	5,909	3,698	4,412	2,147	10,321	5,845
31 Napa	3,352	2,413	3,006	1,589	6,358	4,002
32 Kings	6,407	4,403	6,129	2,985	12,536	7,388
33 Madera	6,023	3,980	5,227	2,440	11,250	6,420
34 Monterey	3,805	2,564	4,027	1,903	7,832	4,467
35 Humboldt	2,810	1,729	2,149	987	4,959	2,716
36 Nevada	2,364	1,661	2,789	1,412	5,153	3,073
37 Mendocino	3,132	1,848	2,919	1,118	6,051	2,966
38 Sutter	4,459	3,147	5,206	2,713	9,665	5,860
39 Yuba	5,302	3,604	4,977	2,515	10,279	6,119
40 Lake	3,597	2,195	4,309	1,754	7,906	3,949
41 San Benito	4,969	3,451	5,101	2,852	10,070	6,303
42 Tehama, etc.	3,513	2,254	4,243	1,910	7,756	4,164
43 Del Norte, etc.	2,945	1,711	2,200	823	5,145	2,534
44 Tuolumne, etc.	2,701	1,767	2,487	1,169	5,188	2,936

Source: UCLA Center for Health Policy Research, 2021-2022 California Health Interview Survey.