



california  
health  
interview  
survey

## CHIS 2021 Methodology Report Series

# Report 1

# Sample Design

September 2022

**CALIFORNIA HEALTH INTERVIEW SURVEY**

**CHIS 2021 METHODOLOGY SERIES**

**REPORT 1**

**SAMPLE DESIGN**

**SEPTEMBER 2022**

*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](http://www.chis.ucla.edu)

This report provides analysts with information about the sampling methods used for CHIS 2021, 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 Methodology Series: Report 1 - Sample Design*. Los Angeles, CA: UCLA Center for Health Policy Research, 2022.

Copyright © 2022 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 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 Control 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, and San Diego County Health and Human Services Agency.

## PREFACE

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

The methodological reports for CHIS 2021 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. An appropriate sample design is a feature of a successful survey, and CHIS 2021 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 data;
- To document data collection procedures so that future iterations of CHIS, or other similar surveys, can replicate those procedures if desired;
- To describe lessons learned from the data collection experience and make recommendations for improving future surveys; and
- To evaluate the level of effort required for the various kinds of data collection undertaken.

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

**Table of Contents**

- 1. CHIS 2021 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 ..... 1-6
  - 1.5 Response Rates ..... 1-12
  - 1.6 Weighting the Sample..... 1-13
  - 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..... 2-5
    - 2.3.1. San Diego Oversample ..... 2-5
    - 2.3.2 Cedars-Sinai Oversample ..... 2-5
    - 2.3.3 Prepaid Cell Oversample ..... 2-5
    - 2.3.4. American Indian and Alaskan Native (AIAN) Oversample ..... 2-6
- 3. SAMPLING HOUSEHOLDS..... 3-1
  - 3.1 Population of Interest..... 3-1
  - 3.2 Analytic Objectives ..... 3-1
  - 3.3 Sample Design ..... 3-2
    - 3.3.1 Supplemental Samples for San Diego County..... 3-5
  - 3.4 Sample Selection and Sample Releases..... 3-8
- 4. WITHIN-HOUSEHOLD SAMPLING ..... 4-1
  - 4.1 Adult Sampling ..... 4-1
  - 4.2 Child Sampling ..... 4-2
    - 4.2.1 Child-First Procedure..... 4-2
    - 4.2.2 Child-then-Adult Ordering ..... 4-3
  - 4.3 Adolescent Sampling ..... 4-3
- 5. ACHIEVED SAMPLE SIZES ..... 5-1
- 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 sample design.....	1-5
Table 1-2. Number of completed interviews by mode of interview and instrument across all samples <sup>2</sup> ...	1-6
Table 1-3. CHIS 2021 survey topic areas by instrument .....	1-8
Table 1-3. CHIS 2021 survey topic areas by instrument (continued) .....	1-9
Table 1-3. CHIS 2021 survey topic areas by instrument (continued) .....	1-10
Table 1-3. CHIS 2021 survey topic areas by instrument (continued) .....	1-11
Table 1-4a. CHIS response rates - Conditional.....	1-12
Table 1-4b. CHIS response rates - Unconditional.....	1-12
Table 2-1: CHIS 2021 Sample Stratification Predictors by Targeted Attribute.....	2-2
Table 2-1: CHIS 2021 Sample Stratification Predictors by Targeted Attribute (continued) .....	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
Table 3-3. Los Angeles County Health Districts .....	3-4
Table 3-4. Multi-county strata.....	3-5
Table 3-5. Initial and final 2021 targets for completed adult interviews by geographic strata (including supplemental samples).....	3-6
Table 3-6. Initial and final 2021 targets for completed adult interviews by geographic strata (excluding supplemental samples) (continued).....	3-7
Table 3-6. Total sample generated and fielded by modeled strata .....	3-8
Table 3-7. Sample release by wave.....	3-9
Table 4-1. Distribution of households with children by child selection probability and year .....	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 .....	5-2
Table 5-2. Number of completed adult interviews by self-reported stratum (continued) .....	5-3
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 stratum.....	5-5
Table 5-5. Number of completed interviews by interview combinations and year.....	5-6
Table 6-1. Complete adult interviews by modeled variables (counts) .....	6-2
Table 6-2. Completed adult interviews by modeled variables (percentages).....	6-3
Table A-1. Design strata definitions for CHIS 2001, 2003, 2005, 2007, 2009, 2011-2012, 2013-2014, 2015-2016, 2017-2018, 2019-2020, and 2021 .....	9-2
Table A-1. Design strata definitions for CHIS 2001, 2003, 2005, 2007, 2009, 2011-2012, 2013-2014, 2015-2016, 2017-2018, 2019-2020, 2021 (continued) .....	9-3
Table A-2. Number of sample pieces selected by design stratum.....	9-4
Table A-3. Number of sample pieces selected by design stratum (continued) .....	9-5

# 1. CHIS 2021 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.

- Report 1 – Sample Design;
- Report 2 – Data Collection Methods;
- Report 3 – Data Processing Procedures;
- Report 4 – Response Rates; and
- Report 5 – Weighting and Variance Estimation.

For further information on CHIS data and the methods used in the survey, visit the California Health Interview Survey Web site at <http://www.chis.ucla.edu> or contact CHIS at [CHIS@ucla.edu](mailto:CHIS@ucla.edu). For methodology reports from previous CHIS cycles, go to <http://www.chis.ucla.edu/chis/design/Pages/methodology.aspx>

The CHIS is a population-based multimode (web and telephone) survey of California’s residential, noninstitutionalized population conducted every other year since 2001 and continually beginning in 2011. CHIS is the nation’s largest state-level health survey and one of the largest health surveys in the nation. The UCLA Center for Health Policy Research (UCLA-CHPR) conducts CHIS in collaboration with multiple funding sources from public, private, and non-profit organizations. CHIS collects extensive information for all age groups on health status, health conditions, health-related behaviors, health insurance coverage, access to health care services, and other health and health-related issues.

The sample is designed and optimized to meet two objectives:

- 1) Provide estimates for large- and medium-sized counties in the state, and for groups of the smallest counties (based on population size), and
- 2) Provide statewide estimates for California’s overall population, its major racial and ethnic groups, as well as several racial and ethnic subgroups.

The CHIS sample is representative of California’s non-institutionalized population living in households. CHIS data and results are used extensively by federal and State agencies, local public health agencies and organizations, advocacy and community organizations, other local agencies, hospitals, community clinics, health plans, foundations, and researchers. These data are used for analyses and



publications to assess public health and health care needs, to develop and advocate policies to meet those needs, and to plan and budget health care coverage and services. Many researchers throughout California and the nation use CHIS data files to further their understanding of a wide range of health related issues (visit UCLA-CHPR's publication page at <http://healthpolicy.ucla.edu/publications/Pages/default.aspx> for examples of CHIS studies).

## **1.2 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, 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 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 an average of 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.

In addition, two oversamples were included. 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 and Asians and aged 50 and older. 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.

In order to provide CHIS data users with more complete and up-to-date information to facilitate analyses of CHIS data, additional information on how to use the CHIS sampling weights, including sample statistical code, is available at <http://www.chis.ucla.edu/chis/analyze/Pages/sample-code.aspx>.

Additional documentation on constructing the CHIS sampling weights is available in the *CHIS*

*2021 Methodology Series: Report 5—Weighting and Variance Estimation* posted at <http://www.chis.ucla.edu/chis/design/Pages/methodology.aspx>. Other helpful information for understanding the CHIS sample design and data collection processing can be found in the four other methodology reports for each CHIS cycle and year.

### **1.3 Sample Design Objectives**

The CHIS 2021 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 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 was stratified into different strata that had higher incidences of individuals with targeted characteristics. For CHIS 2021, 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 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, the following strata were created:

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

This stratification scheme was designed to make use of the most effective predictive variables to target key demographic subgroups in an efficient way that minimizes the impact of the disproportionate sampling on the design effect. Those models that were not sufficiently predictive to add value were excluded. 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.

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.

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

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

Source: UCLA Center for Health Policy Research, 2021 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, we utilized a Prepaid cell oversample and targeted 500 completes to obtain additional in-language 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, 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 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 by mode of interview. Note that these figures were accurate as of data collection completion for 2021 and may differ slightly from numbers in the data files due to data cleaning and edits. Sample sizes to compare against data files you are using are found online at

<http://www.chis.ucla.edu/chis/design/Pages/sample.aspx>.

Table 1-2. Number of completed interviews by mode of interview and instrument across all samples<sup>2</sup>

	Adult	Child	Adolescent
Totals <sup>1</sup>	25,347	4,110	1,192
Completes by Web	22,939	3,822	1,088
Completes by phone	2,408	288	104

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

<sup>1</sup> Includes interviews meeting the criteria as partially complete.

<sup>2</sup> Includes all samples – Main CHIS, prepaid cell phone, Cedar Sinai, and AIAN

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 31 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 34 minutes for the adult, child, and adolescent interviews respectively; the non-English CAWI interviews had an average length of about 60 minutes, 18 minutes, and 30 minutes for the adult, child, and adolescent interviews, respectively. Nearly 8 percent of the adult interviews were completed in a language other than English, as were about 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.

Table 1-3. CHIS 2021 survey topic areas by instrument

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

(continued)

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

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

(continued)



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

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

(continued)

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

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

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

## 1.5 Response Rates

The overall response rates for CHIS 2021 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, the overall household response rate was 9.5 percent (the product of the screener response rate of 13.7 percent and the extended interview response rate at the household level of 69.6 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 Methodology Series: Report 4 – Response Rates*).

The extended interview response rate for the ABS sample varied across the adult (65.1 percent), child (85.3 percent) and adolescent (29.1 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 (see Table 1-4b).

Table 1-4a. CHIS response rates - Conditional

Type of Sample	Screener <sup>1</sup>	Household (given screened) <sup>1</sup>	Adult (given screened) <sup>1</sup>	Child (given screened & eligibility) <sup>1</sup>	Adolescent (given screened & permission) <sup>1</sup>
Overall	13.7%	69.6%	65.1%	85.3%	29.1%

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

<sup>1</sup> The prepaid cell, Cedars-Sinai and AIAN oversamples are not included in these rates.

Table 1-4b. CHIS response rates - Unconditional

Type of Sample	Screener <sup>1</sup>	Household (given screened) <sup>1</sup>	Adult (given screened) <sup>1</sup>	Child (given screened & eligibility) <sup>1</sup>	Adolescent (given screened & permission) <sup>1</sup>
Overall	13.7%	9.5%	8.9%	11.7%	4.0%

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

<sup>1</sup> The prepaid cell, Cedars-Sinai 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, either a spouse/partner or adult child completed a proxy interview for twelve adults. A reduced questionnaire, with questions identified as appropriate for a proxy respondent, was administered.

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

## **1.6 Weighting the Sample**

To produce population estimates from CHIS data, weights were applied to the sample data to compensate for the probability of selection and a variety of other factors, some directly resulting from the design and administration of the survey. The sample was weighted to represent the noninstitutionalized population for each sampling stratum and statewide. The weighting procedures used for CHIS 2021 accomplish the following objectives:

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

As part of the weighting process, a household weight was created for all households that completed the screener interview. This household weight is the product of the “base weight” (the inverse of the probability of selection of the address) and several adjustment factors. The household weight was used to compute a person-level weight, which includes adjustments for the within-household sampling of

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

Population control totals of the number of persons by age, race, and sex at the stratum level for CHIS 2021 were created primarily from the California Department of Finance’s (DOF) 2021 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 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 data with estimates using data from CHIS cycles before 2010. The most accurate California population figures are available when the U.S. Census Bureau conducts the decennial census. For periods between each census, population-based surveys like CHIS must use population projections based on the decennial count. For example, population control totals for CHIS 2009 were based on 2009 DOF estimates and projections, which were based on Census 2000 counts with adjustments for demographic changes within the state between 2000 and 2009. These estimates become less accurate and more dependent on the models underlying the adjustments over time. Using the most recent Census population count information to create control totals for weighting produces the most statistically accurate population estimates for the current cycle, but it may produce unexpected increases or decreases in some survey estimates when comparing survey cycles that use 2000 Census-based information and 2010 Census-based information.

## **1.7 Imputation Methods**

Missing values in the CHIS data files were replaced through imputation for nearly every variable. This was a substantial task designed to enhance the analytic utility of the files. SSRS imputed missing values for those variables used in the weighting process and UCLA-CHPR staff imputed values for nearly every other variable.

Three different imputation procedures were used by SSRS to fill in missing responses for items essential for weighting the data. The first imputation technique was a completely random selection from the observed distribution of respondents. This method was used only for a few variables when the percentage of the items missing was very small. The second technique was hot-deck imputation. The hot-deck approach is one of the most commonly used methods for assigning values for missing responses. Using a hot deck, a value reported by a respondent for a specific item was assigned or donated to a “similar” person who did not respond to that item. The characteristics defining “similar” vary for different variables. To carry out hot-deck imputation, the respondents who answered a survey item formed a pool of donors, while the item non-respondents formed a group of recipients. A recipient was matched to the subset pool of donors based on household and individual characteristics. A value for the recipient was then randomly imputed from one of the donors in the pool. SSRS used hot-deck imputation to impute the same items that have been imputed in all CHIS cycles since 2003 (i.e., race, ethnicity, home ownership, and education). The last technique was external data assignment. This method was used for geocoding variables such as strata, Los Angeles SPA, San Diego HSSA region, and zip where the respondent provided inconsistent information. For such cases geocoding information was used for imputation.

UCLA-CHPR imputed missing values for nearly every variable in the data files other than those imputed by SSRS and some sensitive variables for which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2021 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 can be summarized as a stratified address-based sampling (ABS) design with strategic oversamples of households predicted to have certain attributes. This strategic oversampling is discussed in the Section 2.2. The strata are consistent with prior years, and are defined by county with sub-county substrata for Los Angeles and San Diego counties, as summarized in Table 3-2.

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

The ABS sample is generated from the U.S. Postal Service's Computerized Delivery Sequence File (CDS) which includes all delivery point addresses services by the USPS. CDS provides near-complete coverage of the household population in the United States (e.g., Iannacchione, 2011; Shook-Sa, 2014; Harter et al., 2016). When drawing sample for CHIS, only records flagged as residential or mostly residential are included, as well as P.O. boxes defined as the only way a household can get mail (OWGM, that is, the homeowner has requested no mail delivery at the actual household, just the P.O. Box). Excluded are other P.O. boxes, along with seasonal and vacant households. The study does not cover institutionalized residences/group quarters (e.g., prisons, psychiatric hospitals, long-time care facilities, etc.). The ABS sample for CHIS 2021 was selected via probability sampling methods, and supplied by Marketing Systems Group (MSG).

### **2.2. Targeting Demographics through Predictive Modeling**

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

For CHIS 2021, 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, we also investigated the utility of individual sample flags provided by MSG database information, including the surname flags, child indicator variables, and resident age information as well as PDB block-group characteristics including the density of households with African American residents and households with limited English proficiency.

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

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

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 Interview	Predictive model & geographic density indicator	Predictive model & geographic density indicator
Hispanic	Surname flag, predictive model, & geographic density indicator	Surname flag
African American	Predictive models & geographic density indicator	Geographic density indicator
Native Hawaiian, other Pacific Islander, American Indian, or Alaskan Native	Predictive model	Not specifically targeted

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

Targeted Attribute	Predictor Evaluated	Use in CHIS 2021-2022
No H.S. Diploma	Predictive model	Not specifically targeted
Low income (under \$35,000)	Predictive model	Not specifically targeted
Non-citizen	Predictive model	Not specifically targeted
Presence of children (under 19)	Sample frame flag & predictive model	Predictive model
Presence of adult aged 18-29	Sample frame flag & predictive model	Not specifically targeted
Presence of adult aged 65+	Sample frame flag & predictive model	Sample frame flag & predictive model

These identifying indicators were then used to create a hierarchical set of sample strata and each address in the selected sample was assigned to one of the strata with preference given to the higher ranked strata (for example, a household predicted to be Vietnamese was included in the Vietnamese strata regardless of which other categories they were predicted to be in). For CHIS 2021, the following strata are created:

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

This stratification scheme was designed to make use of the most effective predictive variables to target key demographic subgroups in an efficient way that minimizes the impact of the disproportionate sampling on the design effect. Those models that were not sufficiently predictive to add value were

excluded<sup>1</sup>. 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 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 are 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 generate 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 is selected, random subsamples within modelled 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 only a relative sampling fraction of sample from other strata were selected relative to that in the African American stratum.

---

<sup>1</sup> 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.

## **2.3. Oversampling**

The CHIS design regularly includes additional sample 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. The oversample targeted specific overall quotas by Health and Human Services Agency (HHS) regions, for a total of 112 additional interviews in each region.

### **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 was not included in the publicly released CHIS 2021 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 the ABS frame in CHIS 2021, we utilized a Prepaid cell oversample and targeted 500 completes for this oversample. In particular, this sample was targeted to reach in-language interviews, Hispanic and African American samples, and young adults. We used MSG listed cell sample and then further

screened the sample to only retain interviews from in-language, Hispanic, African-American, and respondents aged 18-24.

We released 102,989 sample pieces to meet our target for this oversample.

#### **2.3.4. American Indian and Alaskan Native (AIAN) Oversample**

CHIS 2021 sought to conduct an additional oversample of 125 adults 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 120 Block Groups that exceed the 60th percentile in AIAN density and also meet the 75% rural definition. The sampled addresses were compared against the main CHIS sample and duplicates were purged.

We released 26,748 sample pieces to meet our target for this oversample.

### 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 in Table 3-1.

Overall, CHIS 2021 was originally designed to yield 20,000 completed adult interviews. The targeted number of adolescent and child (proxy) interviews were established per projections from CHIS 2019-2020. Targets by design strata and for the supplemental samples are discussed in detail in Section 3.3.

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

Characteristics	Interviews (n)
State-wide, Main Study, Overall	
Adults (overall)	20,000
Adolescents (overall) <sup>1</sup>	675
Children (overall) <sup>1</sup>	2,450
Supplemental samples (adults only):	
San Diego County	672
Cedars Sinai	800
Prepaid Cell	500
AIAN	125

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

<sup>1</sup> 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 can be summarized as a stratified ABS design with strategic oversamples of households predicted to have certain attributes. As with previous waves, CHIS 2021 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 the previous wave, CHIS 2021 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 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 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) <sup>1</sup>	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) <sup>2</sup>	32 – Kings
2.1 – North Coastal	33 – Madera
2.2 – North Central	34 – Monterey
2.3 – Central	35 – Humboldt
2.4 – South	36 – Nevada
2.5 – East	37 – Mendocino
2.6 – North Inland	38 – Sutter
3 – Orange	39 – Yuba
4 – Santa Clara	40 – Lake
5 – San Bernardino	41 – San Benito
6 – Riverside	42 – Colusa, Glenn, Tehama
7 – Alameda	43 – Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity
8 – Sacramento	44 – Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne
9 – Contra Costa	
10 – Fresno	
11 – San Francisco	
12 – Ventura	
13 – San Mateo	
14 – Kern	
15 – San Joaquin	
16 – Sonoma	
17 – Stanislaus	
18 – Santa Barbara	
19 – Solano	
20 – Tulare	
21 – Santa Cruz	
22 – Marin	

---

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

<sup>1</sup> Service Planning Areas (SPAs) are analytically important substrata of Los Angeles County.

<sup>2</sup> Health and Human Service Agency (HHSA) regions are analytically important substrata of San Diego County.



Table 3-3. Los Angeles County Health Districts

---

SPA 1 – Antelope Valley

Antelope Valley

SPA 2 – San Fernando Valley

East Valley

Glendale

San Fernando

West Valley

SPA 3 – San Gabriel Valley

Alhambra

El Monte

Foothill

Pasadena

Pomona

SPA 4 – Metro

Central

Hollywood-Wilshire

Northeast

SPA 5 – West

West

SPA 6 – South

Compton

South

Southeast

Southwest

SPA 7 – East

Bellflower

East LA

San Antonio

Whittier

SPA 8 – South Bay

Harbor

Inglewood

Long Beach

Torrance

---

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

Table 3-4. Multi-county strata

---

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

---

Source: UCLA Center for Health Policy Research, 2021 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 112 interviews in each of its six Health and Human Services Agency (HHSA) regions. This resulted in a target of 362 completes in each HHSA region, for a total target of 2,172 completes in San Diego. In the tables below, we show targeted number of adult interviews by geographic strata along with relative population size in California.

The revised 2021 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 and the AIAN oversample.

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

Stratum	Initial Total <sup>1,2</sup>	Oversamples	Final Total <sup>1,2</sup>	Population size <sup>3</sup>
State Total	20,004		20,676	
1 Los Angeles (total) <sup>1</sup>	3,960		3,960	Over 10 million
1.1 – Antelope Valley	250		250	
1.2 – San Fernando Valley	835		835	
1.3 – San Gabriel Valley	686		686	
1.4 – Metro	439		439	
1.5 – West	250		250	
1.6 – South	394		394	
1.7 – East	508		508	
1.8 – South Bay	599		599	
2 San Diego (total) <sup>2</sup>	1,500	672	2,172	3.2 million or greater
2.1 – North Coastal	250	112	362	
2.2 – North Central	250	112	362	
2.3 – Central	250	112	362	
2.4 – South	250	112	362	
2.5 – East	250	112	362	
2.6 – North Inland	250	112	362	
3 Orange	1,230		1,230	900,000 to
4 Santa Clara	762		762	3.2 million
5 San Bernardino	764		764	
6 Riverside	866		866	
7 Alameda	685		685	
8 Sacramento	637		637	
9 Contra Costa	465		465	
10 Fresno	366		366	
11 San Francisco	437		437	600,000 to
12 Ventura	325		325	900,000
13 San Mateo	315		315	
14 Kern	321		321	
15 San Joaquin	271		271	
16 Sonoma	250		250	Medium
17 Stanislaus	250		250	counties
18 Santa Barbara	250		250	100,000 to
19 Solano	250		250	600,000
20 Tulare	250		250	
21 Santa Cruz	250		250	
22 Marin	250		250	

(continued)

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

Stratum	Initial Total <sup>1,2b</sup>	Oversamples	Final Total <sup>1,2</sup>	Population size <sup>3</sup>
23 San Luis Obispo	250		250	
24 Placer	250		250	
25 Merced	250		250	
26 Butte	250		250	
27 Shasta	250		250	
28 Yolo	250		250	
29 El Dorado	250		250	
30 Imperial	250		250	
31 Napa	250		250	
32 Kings	250		250	
33 Madera	250		250	
34 Monterey	250		250	
35 Humboldt	250		250	
36 Nevada	250		250	Small counties
37 Mendocino	250		250	Less than
38 Sutter	250		250	100,000
39 Yuba	250		250	
40 Lake	250		250	
41 San Benito	250		250	
42 Tehama, etc.	200		200	Small counties
43 Del Norte, etc.	200		200	combined
44 Amador, etc.	200		200	
Cedars-Sinai OS		800	800	
Prepaid Cell OS		500	500	
AIAN OS		125	125	

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

<sup>1</sup> 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.

<sup>2</sup> 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.

<sup>3</sup> Based on 2020 California Department of Finance population estimates excluding group quarters.

### 3.4 Sample Selection and Sample Releases

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

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

	Stratum	Modeled Households	Sample Mailed	Mailed Sample with Phone Appended
1.	Vietnamese	12,148	8,306	5,922
2.	Korean	20,314	10,259	6,930
3.	Likely Asian-language Interview	60,231	22,501	16,206
4.	Likely Spanish-language interview	198,188	102,786	74,529
5.	Hispanic	52,589	10,483	6,336
6.	Other high-density non-English	251,555	67,187	44,013
7.	Other Asian	17,380	3,631	2,798
8.	High density African American	21,517	18,938	13,822
9.	HH with children	132,443	45,127	34,439
10.	Other 65+	124,458	9,283	8,092
11.	Residual - Match	64,727	12,124	5,468
12.	Residual – No match	48,824	8,706	5,985

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

The sample for CHIS 2021 was generated monthly and released in 25 waves. Each generation of sample in CHIS 2021 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 of the 25 waves.

Table 3-7. Sample release by wave

Mail Wave	Sample Size	Initial Mailing
1	13,860	3/15/2021
2	13,858	3/15/2021
3	13,862	3/18/2021
4	13,859	3/25/2021
5	13,862	4/1/2021
6	13,861	4/8/2021
7	13,862	4/15/2021
8	13,859	4/22/2021
9	13,862	4/29/2021
10	13,861	5/6/2021
11	13,861	5/13/2021
12	13,863	5/20/2021
13	13,862	5/27/2021
14	13,860	6/3/2021
15	13,862	6/10/2021
16	13,861	6/17/2021
17	13,111	6/24/2021
18	12,606	7/1/2021
19	12,281	7/8/2021
20	11,973	7/15/2021
21	12,491	7/22/2021
22	12,490	7/29/2021
23	7,744	8/12/2021
24	4,944	8/19/2021
25	9,916	9/2/2021

Source: UCLA Center for Health Policy Research, 2021 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, and also describes the differential sampling across two child age groups. Section 4.3 provides a discussion of procedures for choosing one adolescent for interview.

### 4.1 Adult Sampling

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

- The program asks the screener respondent for the number of adults in the household.
- If only one adult lives in the household, then that adult is selected for CHIS.
- If two or more adults live in the household, respondents are asked whether they are the person with the next birthday. If so, they are chosen as the adult respondent. If not, the web program informs the respondent that the adult with the next birthday needs to complete that portion of the survey<sup>2</sup>. 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.

---

<sup>2</sup> The verification question was adapted from Olson & Smyth (2017) to help improve selection accuracy by providing the respondent an active task. CHIS ABS pilot tests experimentally tested the verification question against alternative within-household selection approaches and found it had significantly improved selection accuracy (Wells et al., 2018, 2019).

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

## 4.2 Child Sampling

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

As with previous CHIS waves, children 0-5 years of the selected adult were sampled at twice the rate as older children 6-11 years to increase their representation in the sample. The probability of selecting a child in the 0-5 year group was defined as  $2n_1/(2n_1 + n_2)$ , where  $n_1$  was the number of eligible children ages 0-5 years and  $n_2$  was the number of children ages 6-11 years within the household. The corresponding selection probability for eligible children ages 6-11 years was  $n_2/(2n_1 + n_2)$ . Either the screener respondent or the sampled adult completed the “child interview” about the sampled child. Table 4-1 shows the distribution of households by child age category for CHIS 2021.

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

Child selection probability	Age category of children in household <sup>1</sup>	CHIS 2021	
		n	pct
Equal	Only children 0 to 5 years	1,628	31%
	Only children 6 to 11 years	2,443	47%
Unequal	Children 0 to 5 and 6 to 11 years	1,102	21%
	<b>Total</b>	<b>5,173</b>	<b>100%</b>

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

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

<sup>1</sup> 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 we discontinued this process since with the revised child-adult interview ordering 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 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. 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 Methodology Series: Report 4 – Response Rates*.

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

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

Sample type/interview type	Completed interviews by year		Targets	
	2021	n	pct <sup>2</sup>	
Adult	23,816 <sup>1</sup>	20,676	115.2%	
Child	3,931			
Teen	1159			
Cedars Sinai OS	894 <sup>1</sup>	800	111.8%	
Prepaid Cell OS	503 <sup>1</sup>	500	100.6%	
AIAN OS	134 <sup>1</sup>	125	107.2%	

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

Note: n = sample size; pct = unweighted percent

<sup>1</sup> Includes interviews meeting the criteria as partially complete.

<sup>2</sup> 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. Note that the stratum information reported here corresponds to the design strata but is based on the location of the household as reported by the screener respondent.

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

Reported stratum	CHIS 2021		
	Completes <sup>1</sup>	Target	% of Target
State-wide	23,816	20,676	115.19
1 Los Angeles	4,696	3,981	117.96
2 San Diego	2,523	2,172	116.16
3 Orange	1,374	1,244	110.45
4 Santa Clara	921	762	120.87
5 San Bernardino	850	758	112.14
6 Riverside	956	862	110.90
7 Alameda	898	681	131.86
8 Sacramento	729	626	116.45
9 Contra Costa	533	463	115.12
10 Fresno	402	362	111.05
11 San Francisco	558	434	128.57
12 Ventura	342	327	104.59
13 San Mateo	370	317	116.72
14 Kern	335	318	105.35
15 San Joaquin	291	268	108.58
16 Sonoma	266	250	106.40
17 Stanislaus	261	250	104.40
18 Santa Barbara	293	250	117.20
19 Solano	291	250	116.40
20 Tulare	283	250	113.20
21 Santa Cruz	275	250	110.00
22 Marin	321	250	128.40

(continued)

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

Reported stratum	CHIS 2021		
	Completes <sup>1</sup>	Target	% of Target
23 San Luis Obispo	307	250	122.80
24 Placer	275	250	110.00
25 Merced	283	250	113.20
26 Butte	280	250	112.00
27 Shasta	275	250	110.00
28 Yolo	341	250	136.40
29 El Dorado	272	250	108.80
30 Imperial	315	250	126.00
31 Napa	269	250	107.60
32 Kings	310	250	124.00
33 Madera	287	250	114.80
34 Monterey	259	250	103.60
35 Humboldt	295	250	118.00
36 Nevada	230	250	92.00
37 Mendocino	291	250	116.40
38 Sutter	252	250	100.80
39 Yuba	265	250	106.00
40 Lake	271	250	108.40
41 San Benito	320	250	128.00
42 Tehama, etc.	208	200	104.00
43 Del Norte, etc.	232	200	116.00
44 Tuolumne, etc.	211	200	105.50

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

<sup>1</sup> Includes interviews meeting the criteria as partially complete.

Table 5-3 and Table 5-4 contain the number of completed child and adolescent interviews distributed by reported geographic stratum. The reported stratum in these tables corresponds to the information provided by the screener respondent and is the same as reported in Table 5-2.

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

Reported stratum	Total	Reported stratum	Total
State-wide	3,931	23 San Luis Obispo	36
1 Los Angeles	779	24 Placer	56
2 San Diego	427	25 Merced	48
3 Orange	213	26 Butte	61
4 Santa Clara	154	27 Shasta	39
5 San Bernardino	167	28 Yolo	58
6 Riverside	184	29 El Dorado	40
7 Alameda	158	30 Imperial	76
8 Sacramento	105	31 Napa	28
9 Contra Costa	75	32 Kings	67
10 Fresno	94	33 Madera	54
11 San Francisco	64	34 Monterey	45
12 Ventura	56	35 Humboldt	36
13 San Mateo	68	36 Nevada	31
14 Kern	57	37 Mendocino	31
15 San Joaquin	53	38 Sutter	36
16 Sonoma	32	39 Yuba	55
17 Stanislaus	52	40 Lake	30
18 Santa Barbara	38	41 San Benito	52
19 Solano	31	42 Tehama, etc.	37
20 Tulare	62	43 Del Norte, etc.	30
21 Santa Cruz	41	44 Tuolumne, etc.	23
22 Marin	52		

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

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

Reported stratum	Total	Reported stratum	Total
State-wide	1,159	23 San Luis Obispo	12
1 Los Angeles	212	24 Placer	16
2 San Diego	117	25 Merced	19
3 Orange	68	26 Butte	14
4 Santa Clara	39	27 Shasta	17
5 San Bernardino	44	28 Yolo	25
6 Riverside	51	29 El Dorado	13
7 Alameda	44	30 Imperial	27
8 Sacramento	31	31 Napa	10
9 Contra Costa	26	32 Kings	19
10 Fresno	27	33 Madera	14
11 San Francisco	13	34 Monterey	11
12 Ventura	16	35 Humboldt	17
13 San Mateo	16	36 Nevada	12
14 Kern	25	37 Mendocino	6
15 San Joaquin	12	38 Sutter	11
16 Sonoma	8	39 Yuba	21
17 Stanislaus	17	40 Lake	11
18 Santa Barbara	17	41 San Benito	17
19 Solano	11	42 Tehama, etc.	9
20 Tulare	14	43 Del Norte, etc.	11
21 Santa Cruz	14	44 Tuolumne, etc.	7
22 Marin	18		

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

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

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

Interview combinations <sup>1,2</sup>	CHIS 2021	
	n	pct
Adult only	20,035	81.0
Adult and child	2,694	10.9
Adult and adolescent	723	2.9
Adult, child, and adolescent	364	1.5
Child only	844	3.4
Adolescent only	43	0.2
Child and adolescent only	29	0.1
<b>Total</b>	<b>24,732</b>	<b>100.0</b>

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

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

<sup>1</sup> Includes completed and partial interviews

<sup>2</sup> 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, 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 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, 431 respondents self-identified as Vietnamese (Table 6-1). Of the adult completed interviews not in that group, 120 respondents self-identified as Vietnamese. Additionally, the total number of adult completed interviews from the modeled Vietnamese stratum was 770, while the total number of adult completed interviews not in that stratum was 23,046, 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 22.5% 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 30.4%, which is an improvement over the incidence of the presence of a child or adolescent among adult respondents outside of that group (10.8%). Though the magnitude of improvement was the lowest among this modeled set, this modeling effort still proved effective in reaching the targeted group.



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

Model-Defined Group	Self-Reported Demographics							Completed Adult Interviews <sup>1</sup>
	Vietnamese	Korean	Other Asian	Hispanic / Spanish-Language Household	African American	Age 65+	Child / Adolescent Present in Household	
Vietnamese Household	431	31	228	33	5	165	184	770
Korean Household	98	323	785	48	19	372	341	1,438
Asian-Language Household	455	317	1,638	344	83	967	839	3,722
Spanish-Language Household	30	23	402	3,744	149	1,030	1,436	5,272
Hispanic Household	16	17	312	3,824	91	993	1,395	4,906
Other non-English-Language Household	358	297	1,885	3,700	560	3,368	2,373	11,120
Other Asian Household	466	355	1,942	550	105	1,018	1,012	4,210
African American Household	48	33	325	735	571	946	586	2,961
Household with child (under 19)	397	267	1,999	4,101	511	1,437	4,053	13,349
Age 65+ Household	132	103	850	1,351	381	5,344	409	7,094
No auxiliary data group	2	4	44	94	26	217	113	652
Residual group	2	3	31	95	19	119	95	753
Completed Adult Interviews <sup>1</sup>	551	463	3,281	6,073	1,200	6,961	5,182	23,816

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

<sup>1</sup> Includes interviews meeting the criteria as partially complete.

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

Model-Defined Group	Self-Reported Demographics							Completed Adult Interviews <sup>1</sup>
	Vietnamese	Korean	Other Asian	Hispanic / Spanish-Language Household	African American	Age 65+	Child / Adolescent Present in Household	
Vietnamese Household	56.0%	4.0%	29.6%	4.3%	0.6%	21.4%	23.9%	770
Korean Household	6.8%	22.5%	54.6%	3.3%	1.3%	25.9%	23.7%	1,438
Asian-Language Household	12.0%	8.5%	44.0%	9.2%	2.2%	26.0%	22.5%	3,722
Spanish-Language Household	0.6%	0.4%	7.6%	71.0%	2.8%	19.5%	27.2%	5,272
Hispanic Household	0.3%	0.3%	6.4%	77.9%	1.9%	20.2%	28.4%	4,906
Other non-English-Language Household	3.2%	2.7%	17.0%	33.3%	5.0%	30.3%	21.3%	11,120
Other Asian Household	11.1%	8.4%	46.1%	13.1%	2.5%	24.2%	24.0%	4,210
African American Household	1.6%	1.1%	11.0%	24.8%	19.3%	31.9%	19.8%	2,961
Household with child (under 19)	3.0%	2.0%	15.0%	30.7%	3.8%	10.8%	30.4%	13,349
Age 65+ Household	1.9%	1.5%	12.0%	19.0%	5.4%	75.3%	5.8%	7,094
No auxiliary data group	0.3%	0.6%	6.7%	14.4%	4.0%	33.3%	17.3%	652
Residual group	0.3%	0.4%	4.1%	12.6%	2.5%	15.8%	12.6%	753
Completed Adult Interviews <sup>1</sup>	2.3%	1.9%	13.8%	25.5%	5.0%	29.2%	21.8%	23,816

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

<sup>1</sup> Includes interviews meeting the criteria as partially complete.

## **7. LIMITATIONS FOR CHIS SAMPLE DESIGN**

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

## 8. REFERENCES

- Djangali, A.L., Joseph-David, J., & Trofimovich, L. (2019). *Increasing representativeness through the use of predictive modeling and targeted outreach*. Paper presented at the American Association of Public Opinion Research Conference, Toronto, ON.
- Dutwin, D. (2020). Feedback Loop: Using Surveys to Build and Assess Registration-Based Sample Religious Flags for Survey Research. In *Big Data Meets Survey Science: A Collection of Innovative Methods*. John Wiley & Sons.
- Harter, R., Battaglia, M. P., Buskirk, T. D., Dillman, D. A., English, N., Fahimi, M., Frankel, M. R., Kennel, T., McMichael, J. P., McPhee, C. B., Montaquila, J., Yancey, T., & Zukerberg, A. L. (2016). *Address-based sampling*. Oakbrook Terrace, IL: American Association for Public Opinion Research. <https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx>.
- Iannacchione, V. (2011). The changing role of address-based sampling in survey research. *Public Opinion Quarterly*, 75(3), 556–575.
- McPhee, C., Jackson, M. & Quenneville, G. (2019). *Using model-based stratification to improve sampling efficiency*. Paper presented at the American Association of Public Opinion Research Conference, Toronto, ON.
- Olson, K. & Smyth, J. (2017). Within-household selection in mail surveys: Explicit questions are better than cover letter instructions. *Public Opinion Quarterly*, 81(3), 688-713.
- Shook-Sa, B. E. (2014). Improving the efficiency of address-based sampling frames with the USPS No-Stat File. *Survey Practice*, 7(4). [http://www.surveypractice.org/index.php/SurveyPractice/article/view/264/html\\_8](http://www.surveypractice.org/index.php/SurveyPractice/article/view/264/html_8)
- Wells, B. M., Hughes, T., Park, R., CHIS Redesign Working Group, Rogers, T. B., & Ponce, N. (2018). *Evaluating the California Health Interview Survey of the future: Results from a methodological experiment to test an address-based sampling mail push-to-web data collection*. Los Angeles, CA: UCLA Center for Health Policy Research.

Wells, B. M., Hughes, T., Park, R., CHIS Redesign Working Group, & Ponce, N. (2019). *Evaluating the California Health Interview Survey of the future: Results from a statewide pilot of an address-based sampling mail push-to-web data collection*. Los Angeles, CA: UCLA Center for Health Policy Research.

## **APPENDIX A**

Appendix A contains supplemental information on the CHIS 2021 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 separately by design stratum.

Table A-1. Design strata definitions for CHIS 2001, 2003, 2005, 2007, 2009, 2011-2012, 2013-2014, 2015-2016, 2017-2018, 2019-2020, and 2021

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

(continued)

Table A-1. Design strata definitions for CHIS 2001, 2003, 2005, 2007, 2009, 2011-2012, 2013-2014, 2015-2016, 2017-2018, 2019-2020, 2021 (continued)

County	2015-2016, 2017-2018, 2019-2020, 2021 Strata	2013-2014 Strata	2005, 2007, 2009, 2011-2012 Strata	2001, 2003 Strata
Lake	40	40	40	37
Mendocino	37	37	37	
Sutter	38	38	38	39
Yuba	39	39	39	
Colusa	42	42	42	38
Glenn				
Tehama				
Nevada	36	36	36	40
Humboldt	35	35	35	35
Del Norte				
Lassen				36
Modoc		43		
Plumas	43		43	40
Sierra				
Trinity				36
Siskiyou		43.2		
Amador				
Alpine				
Inyo		44		
Mariposa	44		44	41
Mono				
Tuolumne		44.1		
Calaveras		44.2		

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



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

Sampling stratum	Total	Sample with Appended phones
State-wide	319,331	224,540
1 Los Angeles	70,306	49,407
2 San Diego	31,259	22,364
3 Orange	16,602	12,120
4 Santa Clara	9,732	6,639
5 San Bernardino	13,806	10,162
6 Riverside	14,864	11,304
7 Alameda	8,624	6,153
8 Sacramento	8,041	5,819
9 Contra Costa	5,958	4,549
10 Fresno	7,095	4,873
11 San Francisco	5,468	3,628
12 Ventura	4,119	3,205
13 San Mateo	3,796	2,729
14 Kern	6,029	4,204
15 San Joaquin	4,820	3,531
16 Sonoma	2,813	1,960
17 Stanislaus	4,486	3,221
18 Santa Barbara	3,100	2,152
19 Solano	3,637	2,823
20 Tulare	4,878	3,396
21 Santa Cruz	2,801	1,913
22 Marin	2,856	2,095
23 San Luis Obispo	2,615	1,771
24 Placer	2,935	2,268
25 Merced	5,247	3,502
26 Butte	3,140	2,070
27 Shasta	3,378	2,319
28 Yolo	2,815	1,853
29 El Dorado	2,823	2,085

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

Sampling stratum	Total	Sample with Appended phones
30 Imperial	5,909	3,698
31 Napa	3,352	2,413
32 Kings	6,407	4,403
33 Madera	6,023	3,980
34 Monterey	3,805	2,564
35 Humboldt	2,810	1,729
36 Nevada	2,364	1,661
37 Mendocino	3,132	1,848
38 Sutter	4,459	3,147
39 Yuba	5,302	3,604
40 Lake	3,597	2,195
41 San Benito	4,969	3,451
42 Tehama, etc.	3,513	2,254
43 Del Norte, etc.	2,945	1,711
44 Tuolumne, etc.	2,701	1,767

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