

# California Health Interview Survey

## CHIS 2021 Sample Design

Making  
California's  
Voices  
Heard on  
Health



10960 Wilshire Blvd.  
Suite 1550  
Los Angeles, CA  
90024  
t: 310.794.0909  
f: 310.794.2686  
chis@ucla.edu

### Overview

The California Health Interview Survey (CHIS) began in 2001 as a biennial population-based, omnibus health survey of Californians. It remains the largest survey in California and the largest continuous state health survey in the country. CHIS 2021 is the eleventh data collection cycle following CHIS 2001, 2003, 2005, 2007, 2009, 2011-2012, 2013-2014, 2015-2016, 2017-2018, and 2019-2020.

CHIS became a continuous survey in 2011, with ongoing data collection throughout a two-year cycle. The change to a continuous survey enhances the role of CHIS as a critical and timely source of public health surveillance information for California. While this change significantly streamlines CHIS operations and data production, most aspects of the CHIS project, including the overall sample design, remain the same.

CHIS randomly selects one adult to interview in each randomly sampled and participating household throughout California. In addition to adults, interviews are conducted about the health of children (ages 0 to 11, by proxy with an adult sufficiently knowledgeable about the health of the selected child) and adolescents (age 12 to 17, directly interviewed following parental permission). The selected adult respondent must be the parent or legal guardian of children and/or adolescents living in the household in order for children and/or adolescents to be eligible for participation.

The CHIS 2021 sample design is a geographically stratified address-based sample (ABS) design with strategic oversamples of households predicted to have certain attributes. Under the ABS frame, CHIS 2021 included additional samples for specialized analyses of geographic areas and hard-to-reach populations. To better target populations not adequately covered under the ABS frame, CHIS added a supplemental prepaid cell phone sample to the primary ABS sample. More information about the ABS sample design and oversamples, the prepaid cell phone sample, and the multi-mode data collection approach is provided in the following sections.

To capture the rich diversity of the California population, interviews are conducted in six

languages: English, Spanish, Chinese (Mandarin and Cantonese dialect), Vietnamese, Korean, and Tagalog.

Overall, the CHIS sample is designed to provide population-based estimates for most California counties and all major ethnic groups, including several ethnic subgroups. The sample is designed to meet and optimize two goals:

- To provide local-level estimates for counties with populations of 60,000 or more for local planning and comparisons among counties;
- To provide statewide estimates for California's overall population, its major race/ethnic groups, as well as for several Asian and Latino ethnic groups.

CHIS data and statistical estimates are publicly available to local, state and federal agencies; community-based organizations; health care providers and organizations; advocacy groups; policy makers; and individual citizens. CHIS results and data are available through multiple dissemination channels:

- CHIS website with a free, easy-to-use interactive query system, AskCHIS, for tailored data estimates ([ask.chis.ucla.edu](http://ask.chis.ucla.edu))
- Public-use data files for researchers and analysts ([chis.ucla.edu/chis/data/Pages/public-use-data.aspx](http://chis.ucla.edu/chis/data/Pages/public-use-data.aspx))
- Published reports and scholarly publications
- The UCLA Center for Health Policy Research Data Access Center (for more, detailed data files available only through a secure environment, see: [chis.ucla.edu/chis/data/Pages/confidential.aspx](http://chis.ucla.edu/chis/data/Pages/confidential.aspx))

CHIS provides technical assistance for researchers and all users of web-based and public-use data. For assistance, please email [dacchpr@ucla.edu](mailto:dacchpr@ucla.edu) or call 310-794-6783.

## The CHIS 2021 Sample in Detail

### *Address-based sampling*

Starting in 2019, CHIS transformed from a dual-frame landline/cellphone random-digit dial (RDD) methodology to an address-based sample (ABS) methodology with a multi-mode data collection that takes place on the web or by telephone.

As with CHIS 2019-2020, the ABS sample for CHIS 2021 is generated from the United States Postal Service's (USPS) Computerized Delivery Sequence (CDS) file which includes all delivery point addresses serviced by USPS and has near-complete coverage of the household population of the United States. The ABS sample is geographically stratified by county, groups of small counties, and some sub-county areas (see Table 2 at the end of this document). Only residential households are included in the sample, and the following households and persons are not eligible: 1) addresses outside the state of California; 2) institutionalized residences; 3) group quarters (those with nine or more people unrelated persons).

### *Targeting through predictive modeling*

CHIS 2021 employed Big Data techniques to help target specific demographic groups commonly unrepresented. The process appends auxiliary data (e.g., voter registration databases, consumer databases, surname databases) to prior survey data and uses 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. Individual sample flags and geographic density indicators of households with certain attributes were employed during model evaluation. CHIS specifically targeted Asian households including Korean and Vietnamese, households that are more likely to conduct Asian-language interviews, Latino and Spanish-speaking households, households with African American residents, households with children under 19, and households with adults aged over 65. CHIS then oversampled these strata with the aim of maximizing the different key subgroup incidences while minimizing the negative effects on variance estimation.

### *CHIS 2021 Sample Additions*

CHIS 2021 oversampled 800 Latinos and Asians aged 50 and older in LA County Special Planning Areas (SPAs) 1,2,4, and 5. This oversample is under the ABS frame.

Oversamples of AIAN have been conducted in CHIS 2001, CHIS 2011-2012, and CHIS 2018. In 2021, CHIS conducted an additional oversample of 125 adults who identify as American Indian or Alaska Native (AIAN) and live in rural areas. This oversample is the ABS sample.

CHIS 2021 added a prepaid cell phone frame in the sample

design to better represent populations not adequately covered by the primary ABS frame. The prepaid cell phone sample was targeted to reach respondents who were either aged 18 to 24, Hispanic, African American, or would like to take the survey in one of the non-English languages offered for CHIS 2021. The targeted sample size is 500.

### *Mail push-to-web with a telephone follow-up*

For the ABS sample in CHIS 2021, respondents are invited to participate in CHIS through a series of mail invitations, similar to the advance letter used in previous cycles. Respondents receive an initial invitation letter with a \$2.00 incentive, a multilingual letter in all CHIS languages, and frequently asked questions. This is followed by a reminder pressure sealed postcard, standard letter, and a final postcard. The prominent language of all mail materials was tailored for households likely to be Spanish-speaking or likely to speak one of the four Asian languages offered in CHIS.

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. Each mail invitation included a call-in phone number allowing the household to complete the interview with a trained telephone interviewer. Almost half of the total telephone completes (which account for about ten percent of all CHIS adult completes) were a result of call-ins.

Once an eligible adolescent was identified and permission to recontact was provided from a parent, the first mailing for the adolescent was addressed to the parent and contained an interior envelope with a letter addressed to the adolescent. A reminder letter addressed to the adolescent followed as well as a text reminder and telephone (if a phone number was available). All adolescents were offered a \$10.00 gift card for completing the survey. For adult interviews that did not result in parental permission to interview an eligible adolescent, parents were re-contacted with a unique refusal conversion effort using a similar mailing to the nested mailer which included an additional incentive for the parent.

### *Sample Size by Age Category*

As many as three persons will be sampled from each household, one from each of the three age categories: adults (age 18 and older), adolescents (age 12–17), and children (age 0–11). To increase the representation in the sample, children aged 0-5 were sampled at twice rate as older children 6-11 years. Adults and adolescents answer survey questions for themselves, and child data are collected by proxy interview with an adult respondent.

To maximize the child sample in CHIS 2021, the child rostering section was moved up in the adult survey. If an eligible child was identified, the selected adult proceeded with completing the child interview before resuming the adult survey. This protocol

was a departure from previous cycles where now essentially every child interview was conducted prior to the adult interview.

### Sample Size by County

The CHIS sample is stratified to produce local-level estimates for most counties. California’s 58 counties are arranged into sampling strata based on size and analytic goals to produce 56 sampling strata. Table 1 describes the overall stratification approach for the 58 counties, and Table 2 provides target CHIS sample sizes for the 44 county and county group strata for the ABS sample.

Los Angeles and San Diego counties have 8 and 6 sub-county strata, respectively. Of the remaining 56 counties, 39 comprise single-county strata and the remaining 17 counties are arranged into three multiple-county strata. The CHIS sample is allocated to achieve the project’s goals of providing estimates for as many counties as possible and to provide estimates for major race/ethnic groups and subgroups. The CHIS 2021 sample, as in previous CHIS cycles, includes a minimum sample size of at least 200 households for the smallest counties and county groups in a single year; the largest county will have a sample of

nearly 4,000 households in a single year (see Table 2). The CHIS 2021 sample also had minimum sample targets for the 26 Los Angeles County Health Districts and for each county within the grouped county strata.

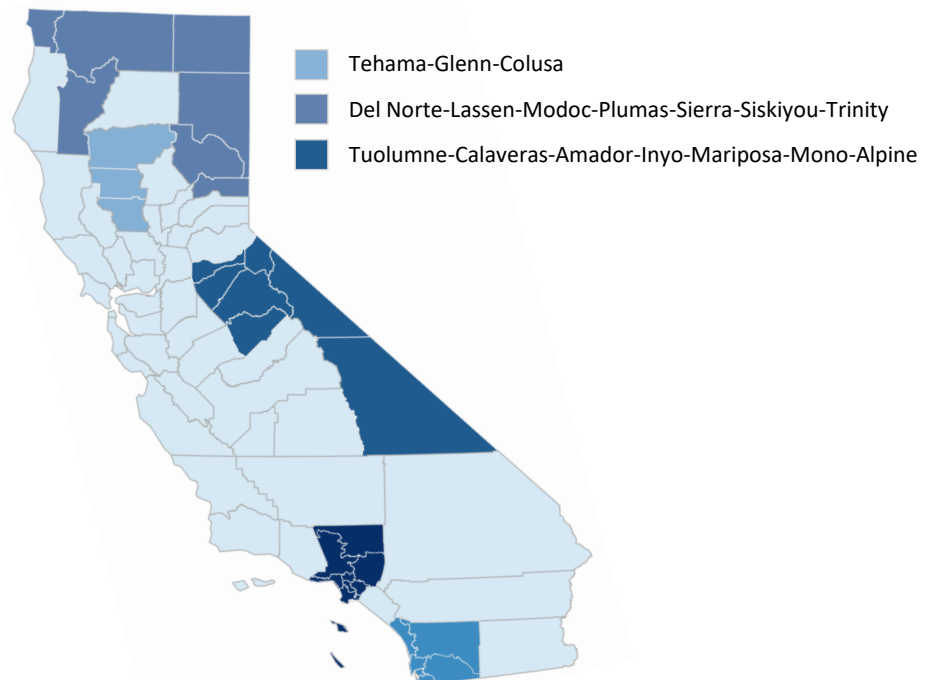
The stratum sample goals in Table 2 are calculated by proportionally allocating the state target using population estimates and projections in 2021 from the California Department of Finance (DOF). In this proportional allocation, a minimum sample size is maintained to ensure sufficient sample for analysis of smaller counties. The figures in Table 2 reflect estimated sample targets based on information available at the time of this report and may be updated in consultation with the CHIS data collection subcontractor as DOF information is updated. Agencies and organizations may partner with CHIS to increase the sample size or add county-specific content in their county through supplemental funding. Oversamples are not included in Table 2 and such oversamples for CHIS 2021 are discussed in [CHIS 2021 Methodology Report 1— Sample Design](#).

For more details regarding the CHIS 2021 sample design, please visit the [CHIS Methodology webpage](#).

**TABLE 1. CHIS 2021 SAMPLING STRATA TYPE AND NUMBER**

STRATUM TYPE	NUMBER OF COUNTIES AND STRATA
Counties with sub-county strata	2 counties, 14 strata total between them
Counties as single stratum	39 counties, 39 strata
Counties combined into single stratum	17 counties, 3 strata total
<b>Total</b>	<b>58 counties, 56 strata</b>

**FIGURE 1. CHIS 2021 SAMPLING STRATA**



# California Health Interview Survey

Making  
California's  
Voices  
Heard on  
Health



10960 Wilshire Blvd.  
Suite 1550  
Los Angeles, CA  
90024  
t: 310.794.0909  
f: 310.794.2686  
chis@ucla.edu

**TABLE 2. CHIS 2021 INITIAL ABS SAMPLE SIZE TARGETS BY COUNTY**

COUNTY NAME	TARGET ADULT SAMPLE
Los Angeles (total)	3,960
San Diego (total)	1,500
Orange	1,230
Santa Clara	762
San Bernardino	764
Riverside	866
Alameda	685
Sacramento	637
Contra Costa	465
Fresno	366
San Francisco	437
Ventura	325
San Mateo	315
Kern	321
San Joaquin	271
Sonoma	250
Stanislaus	250
Santa Barbara	250
Solano	250
Tulare	250
Santa Cruz	250
Marin	250
San Luis Obispo	250
Placer	250
Merced	250
Butte	250
Shasta	250
Yolo	250
El Dorado	250
Imperial	250
Napa	250
Kings	250
Madera	250
Monterey	250
Humboldt	250
Nevada	250
Mendocino	250
Sutter	250
Yuba	250
Lake	250
San Benito	250
Tehama-Glenn-Colusa	200
Del Norte-Lassen-Modoc-Plumas-Sierra-Siskiyou-Trinity	200
Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	200
<b>TOTAL</b>	<b>20,000</b>