



california  
health  
interview  
survey

## CHIS 2017 DATA DICTIONARY

Covered California Data Dictionary

een

October 2018



california  
health  
interview  
survey

# **CHIS 2017 DATA DICTIONARY**

## **Teen Survey**

### **Source**

October 2018

teen.sas7bdat

Contact:  
California Health Interview Survey  
UCLA Center for Health Policy Research  
10960 Wilshire Blvd., Suite 1550  
Los Angeles, CA 90024  
Email: [chis@ucla.edu](mailto:chis@ucla.edu)

***Suggested Citation:***

California Health Interview Survey. CHIS 2017 Teen Survey. UCLA Center for Health Policy Research. Los Angeles, CA: October 2018

# TABLE OF CONTENTS

<b>1. Introduction.....</b>	<b>1</b>
1.1 CHIS 2017 Teen Survey Data File.....	1
1.2 Accompanying Files.....	1
<b>2. What’s New in CHIS 2017.....</b>	<b>2</b>
<b>3. CHIS 2017 Sample Design and Methodology .....</b>	<b>4</b>
1.1 Overview .....	4
1.2 Switch to a Continuous Survey .....	5
1.3 Sample Design Objectives.....	6
1.4 Data Collection.....	9
1.6 Weighting the Sample .....	16
1.7 Imputation Methods.....	17
<b>4. Sample Code for Analysis and Pooling of CHIS Data.....</b>	<b>19</b>
<b>5. Restricted Variables .....</b>	<b>20</b>
<b>6. CHIS Data Dictionary.....</b>	<b>21</b>

**Click on the “Bookmarks” icon to navigate between sections of the data dictionary:**



# 1. Introduction

## 1.1 CHIS 2017 Teen Survey Data File

The 2017 California Health Interview Survey (CHIS 2017) Teen Data File consists of individual records obtained from the 2017 data collection period of the CHIS 2017-2018 Teen survey.

The UCLA-Center for Health Policy Research (UCLA-CHPR) is responsible for maintaining consistent standards to protect respondent confidentiality as specified in approved protocols by the UCLA IRB (IRB# 17-000362) and the California Committee for the Protection of Human Subjects (12-05-0176). Organizations that receive CHIS data are required to complete a data use agreement with UCLA-CHPR. The data files listed below must be maintained in accordance with the provisions of the data use agreement between the California Department of Public Health (CDPH) and UCLA-CHPR. In order to protect respondent confidentiality, UCLA-CHPR maintains the most confidential and sensitive CHIS data only in its files at the Data Access Center (DAC) located at UCLA-CHPR. Researcher access to confidential data, including respondent latitude and longitude, is available by application to the DAC. For more information, please contact [dachpr@ucla.edu](mailto:dachpr@ucla.edu). Limited technical assistance is also available from CHIS – please send email to [chis@ucla.edu](mailto:chis@ucla.edu).

## 1.2 Accompanying Files

In addition to the data file are several accompanying files that facilitate use of the data file; some are not necessary for data analysis but add convenience in utilizing the main data.

- a. **Data file:** TEEN.SAS7BDAT, TEEN.SAV, TEENF.DTA
- b. **Proc format file:** TEEN\_PROC\_FORMAT.SAS
- c. **Format file:** TEEN\_FORMAT.SAS
- d. **Label file:** TEEN\_LABEL.SAS
- e. **Imputation flag file:** TEENF.SAS7BDAT, TEENF.SAV, TEENF.DTA
- f. **Others:** TEEN.XPT, TEENF.XPT

## 2. What's New in CHIS 2017

This document describes new and notable design features and data collected in CHIS 2017. CHIS data users should review the information below and our detailed online documentation before analyzing or reporting CHIS data. Please visit the following page for more documentation on CHIS methods: <http://chis.ucla.edu/chis/design/Pages/methodology.aspx>

### **New and Updated Survey Questions**

Survey questions are added, removed, and modified in each two-year cycle of CHIS to meet stakeholders' needs and monitor emerging public health concerns. Questions are removed from the interview to reduce its length and save data collection costs when topics are no longer relevant for public health surveillance, or when they are not funded by a sponsor. Most CHIS questions remain in the interview across CHIS cycles.

For CHIS 2017 approximately 80 percent of the content continued from CHIS 2016. Occasionally, we make changes to question wording based on methodological evaluations or when user feedback strongly suggests that changes will produce better data. Otherwise, we keep questions consistent across years to aid in trending. Reinstated questions were asked in cycles prior to 2015-2016, and again in 2017-2018. Noteworthy additions to CHIS 2017 include:

#### **New adult interview questions**

- Marijuana use<sup>1</sup>, frequency and methods of use<sup>3</sup>
- Female pregnancy intentions<sup>1</sup>
- Male birth control use<sup>1</sup>
- Voting eligibility and engagement<sup>1</sup>
- Current WIC enrollment<sup>2</sup>, WIC participation in the past 5 years<sup>1</sup>, and satisfaction with WIC<sup>3</sup>
- Prescription drug misuse<sup>2</sup> and heroin use<sup>2</sup>
- Pre-Exposure Prophylaxis<sup>3</sup>
- Reinstated content:
  - Smoking cessation<sup>1</sup>
  - Current birth control use and pregnancy status<sup>1</sup>
  - Race most identified with<sup>1</sup>
  - Paying off mortgage and loan<sup>1</sup> (used in creation of the Elder Index™)
  - Medical debt<sup>1</sup>
  - Usage of health savings accounts<sup>1</sup>

#### **New child interview questions**

- Reclaimed content:
  - Prescription medication and therapy needs<sup>1</sup>
  - Developmental concerns<sup>1</sup>
  - Usage of public dental health programs<sup>1</sup>

#### **New teen interview questions**

- Sleep and technology use<sup>1</sup>
- Marijuana use<sup>2</sup>, frequency of use<sup>2</sup>, and methods of use<sup>3</sup>
- Birth control use and counseling<sup>2</sup>

<sup>1</sup> Available in all CHIS data products including Public Use Files (PUFs)

<sup>2</sup> Available through AskCHIS (<http://ask.chis.ucla.edu>)

<sup>3</sup> Confidential data is only available through the Data Access Center (DAC). For more information please visit <http://healthpolicy.ucla.edu/chis/data/Pages/confidential.aspx>

For a full list of topics covered in CHIS 2017, please visit <http://chis.ucla.edu/chis/design/Pages/survey-topics.aspx>. The full CHIS 2017 questionnaire is available at <http://chis.ucla.edu/chis/design/Pages/questionnairesEnglish.aspx>.

### **Improved Sample Coverage**

In an effort to improve cell phone coverage of California residents, CHIS 2017 included a sample of likely residents (based on zip code) with out-of-state cell phone numbers to better capture recent imports to the state. CHIS 2017 also oversampled residents under 65 to increase the ability to reach households with children and teens.

### **Oversamples**

CHIS 2017 continued to oversample Korean and Vietnamese Americans as has been done in previous cycles since CHIS 2001. For the first-time ever, CHIS was able to expand listed sample of Korean and Vietnamese surnames to include cell phone numbers. CHIS 2017 also saw oversamples in two Southern California counties. San Diego Health and Human Services provided funds to oversample San Diego County while the Imperial County Health Department also provided funds to oversample residents of northern Imperial County.

### **New Vendor**

As the result of a competitive bidding process, SSRS conducted the CHIS 2017-2018 data collection under contract with the UCLA Center for Health Policy Research.

### **Data Collection Timeline**

Like past CHIS data collection, CHIS 2017 data were collected as part of a two-year cycle (2017-2018). CHIS 2017 data were collected between June 2017 and December 2017.

Approximately 50 percent of adult interviews were conducted from landline phone numbers and 50 percent from cell phone numbers.

From CHIS 2011 forward, single-year data are available representing a yearly cross-section of the California's population. Relative to the larger, two-year CHIS data files available prior to 2011, small populations (such as child, teen, or some racial/ethnic groups) or rare conditions and characteristics will have fewer observations in the single-year data file. In such cases, pooling two or more single-year data sets may be required to achieve sufficient sample size and statistical stability.

Users who need more information about pooling or trending data over time should review the *Analyze CHIS Data* website at <http://chis.ucla.edu/chis/analyze/Pages/default.aspx> or go to the *CHIS FAQ* page at <http://chis.ucla.edu/chis/faq/Pages/default.aspx>. You can also email questions to [dacchpr@ucla.edu](mailto:dacchpr@ucla.edu).

### 3. CHIS 2017 Sample Design and Methodology

#### 1.1 Overview

A series of five methodology reports are available with more detail about the methods used in CHIS 2017.

- 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 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 the California Department of Public Health and the California Department of Health Care Services. CHIS collects extensive information for all age groups on health status, health conditions, health-related behaviors, health insurance coverage, access to health care services, and other health and health-related issues.

The sample is designed and optimized to meet two objectives:

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

The CHIS sample is representative of California’s non-institutionalized population living in households. CHIS data and results are used extensively by federal and State agencies, local public health agencies and organizations, advocacy and community organizations, other local agencies, hospitals, community clinics, health plans, foundations, and researchers. These data are used for analyses and publications to assess public health and health care needs, to develop and advocate policies to meet those needs, and to plan and budget health care coverage and services. Many researchers throughout California



and the nation use CHIS data files to further their understanding of a wide range of health related issues (visit UCLA-CHPR's publication page at <http://healthpolicy.ucla.edu/publications/Pages/default.aspx> for examples of CHIS studies).

## 1.2 Switch to a Continuous Survey

From the first CHIS cycle in 2001 through 2009, CHIS data were collected during a 7 to 9 month period every other year. Beginning in 2011, CHIS data have been collected continually over a 2-year cycle. This change was driven by several factors including the ability to track and release information about health in California on a more frequent and timely basis and to eliminate potential seasonality in the biennial data.

CHIS 2017 data were collected between June and December, 2017. As in previous CHIS cycles, weights are included with the data files and are based on the State of California's Department of Finance population estimates and projections, adjusted to remove the population living in group quarters (such as nursing homes, prisons, etc.) and thus not eligible to participate in CHIS. When the weights are applied to the data, the results represent California's residential population during that year for the age group corresponding to the data file in use (adult, adolescent, or child). In CHIS 2017, data users will be able to produce single-year estimates using the weights provided.

**See what's new in the 2017 CHIS sampling and data collection here:**

<http://www.chis.ucla.edu/chis/design/Documents/whats-new-chis-2017.pdf>

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

Additional documentation on constructing the CHIS sampling weights is available in the *CHIS 2017 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 year.

### 1.3 Sample Design Objectives

The CHIS 2017 sample was designed to meet the two sampling objectives discussed above: (1) provide estimates for adults in most counties and in groups of counties with small populations; and (2) provide estimates for California’s overall population, major racial and ethnic groups, and for several smaller racial and ethnic subgroups.

To achieve these objectives, CHIS employed a dual-frame, multi-stage sample design. The random-digit-dial (RDD) sample included telephone numbers assigned to both landline and cellular service. The RDD sample was designed to achieve the required number of completed adult interviews by using approximately 50% landline and 50% cellular phone numbers. For the RDD 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 period—to provide health estimates for adults at the local level. Asian surname sample list frames added 38 Korean, and 74 Vietnamese adult interviews based on self-identified ethnicity for the 2017 survey year.<sup>1</sup> Additional samples from both the landline and cell phone frames produced 635 interviews in 2017 within San Diego County. Furthermore, an address-based sample from the USPS Delivery Sequence File produced 332 landline or cell phone interviews in 2017 within the northern part of Imperial County.

Within each geographic stratum, residential telephone numbers 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, and the adult sufficiently knowledgeable about the child’s health completed the child interview.

---

<sup>1</sup> For the 2017, all sample frames produced totals of 113 Korean, and 148 Vietnamese adult interviews.

Table 1-1. California county and county group strata used in the CHIS 2017 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. Plumas, Sierra, Siskiyou,
4. Santa Clara	24. Placer	Lassen, Modoc, Trinity, Del Norte
5. San Bernardino	25. Merced	44. Mariposa, Mono, Tuolumne,
6. Riverside	26. Butte	Alpine, Amador, Calaveras, Inyo

Source: UCLA Center for Health Policy Research, 2017 California Health Interview Survey.

The CHIS RDD sample is of sufficient size to accomplish the second objective (produce estimates for the state’s major racial/ethnic groups, as well as many ethnic subgroups). However, given the smaller sample sizes of one-year data files, two or more pooled cycles years of CHIS data are generally required to produce statistically stable estimates for small population groups such as racial/ethnic subgroups, children, teens, etc. To increase the precision of estimates for Koreans and Vietnamese, areas with relatively high concentrations of these groups were sampled at higher rates. These geographically targeted oversamples were supplemented by telephone numbers associated with group-specific surnames, drawn from listed telephone directories to increase the sample size further for Koreans and Vietnamese.

To help compensate for the increasing number of households without landline telephone service, a separate RDD sample was drawn of telephone numbers assigned to cellular service. In CHIS 2017, the

goal was to complete approximately 50% of all RDD interviews statewide with adults contacted via cell phone. Because the geographic information available for cell phone numbers is limited and not as precise as that for landlines, cell phone numbers were assigned to the same 44 geographic strata (i.e., 41 strata defined by a single county and 3 strata created by multiple counties) using a classification associated with the rate center linked to the account activation. The cell phone stratification closely resembles that of the landline sample and has the same stratum names, though the cell phone strata represent slightly different geographic areas than the landline strata. The adult owner of the sampled cell phone number was automatically selected for CHIS. Cell numbers used exclusively by children under 18 were considered ineligible. A total of 448 teen interviews and 1,600 child interviews were completed in CHIS 2017 with approximately 51% of teen interviews and 66% of child interviews coming from the cell phone sample.

The cell phone sampling method used in CHIS has evolved significantly since its first implementation in 2007 when only cell numbers belonging to adults in cell-only households were eligible for sampling adults. These changes reflect the rapidly changing nature of cell phone ownership and use in the US.<sup>2</sup> There have been three significant changes to the cell phone sample since 2009. First, all cell phone sample numbers used for non-business purposes by adults living in California were eligible for the extended interview. Thus, adults in households with landlines who had their own cell phones or shared one with another adult household member could have been selected through either the cell or landline sample. The second change was the inclusion of child and adolescent extended interviews. The third, enacted in CHIS 2015-2016 was to increase the fraction of the sample comprised of cell phones from 20% to 50% of completed interviews. In 2017, we additionally sampled out-of-area cell phone numbers. These are cell phone numbers with exchanges outside of California that can be matched to an address that is within California, indicating that the owner of the cell phone resides in California but purchased a cell phone in another state.

The cell phone sample design and targets by stratum of the cell phone sample have also changed throughout the cycles of the survey. In CHIS 2007, a non-overlapping dual-frame design was implemented where cell phone only users were screened and interviewed in the cell phone sample. Beginning in 2009, an overlapping dual-frame design has been implemented. In this design, dual phone users (e.g., those with both cell and landline service) can be selected and interviewed from either the landline or cellphone samples.

The number of strata has also evolved as more information about cell numbers has become available. In CHIS 2007, the cell phone frame was stratified into seven geographic sampling strata created using telephone area codes. In CHIS 2009 and 2011-2012, the number of cell phone strata was

---

<sup>2</sup> <https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201806.pdf>

increased to 28. These strata were created using both area codes and the geographic information assigned to the number. Beginning in CHIS 2011, with the availability of more detailed geographic information, the number of strata was increased to 44 geographic areas that correspond to single and grouped counties similar to the landline strata. The use of 44 geographic strata continued in CHIS 2017.

## **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 dialects), Vietnamese, Korean, and Tagalog. Tagalog interviews were conducted for part of the CHIS 2013-2014 cycle, but 2015-2016 were the first cycle years that Tagalog interviews were conducted from the beginning of data collection. 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 designed the methodology and collected data for CHIS 2017, 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, SSRS staff interviewed one randomly selected adult in each sampled household, and 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. Children and adolescents were generally sampled at the end of the adult interview. If the screener respondent was someone other than the sampled adult, children and adolescents could be sampled as part of the screening interview, and the extended child (and adolescent) interviews could be completed before the adult interview. This “child-first” procedure was first used in CHIS 2005 and has been continued in subsequent CHIS cycles because it substantially increases the yield of child interviews. While numerous subsequent attempts were made to complete the adult interview for child-first cases, the final data contain completed child and adolescent interviews in households for which an adult interview was not completed. Table 1-2 shows the number of completed adult, child, and adolescent interviews in CHIS 2017 by the type of sample (landline RDD, surname list, cell RDD, and ABS). Note that these figures were accurate as of data collection completion for 2017 and may differ slightly from numbers in the data files due to data cleaning and edits. Sample sizes to compare against data files you are using are found online at <http://www.chis.ucla.edu/chis/design/Pages/sample.aspx>.

Table 1-2. Number of completed CHIS 2017 interviews by type of sample and instrument

Type of sample <sup>1</sup>	Adult <sup>2</sup>	Child	Adolescent
Total all samples	21,153	1,600	448
Landline RDD <sup>3</sup>	9,831	501	206
Cell RDD	10,722	1,037	219
Vietnamese surname list landline	47	6	2
Vietnamese surname list cell phone	44	4	3
Korean surname list landline	133	9	3
Korean surname list cell phone	19	1	
Both Korean and Vietnamese Landline	18		
Imperial County ABS Oversample	339	42	15

Source: UCLA Center for Health Policy Research, 2017 California Health Interview Survey.

<sup>1</sup> Completed interviews listed for each sample type refer to the sampling frame from which the phone number was drawn. Interviews could be conducted using numbers sampled from a frame with individuals who did not meet the target criteria for the frame but were otherwise eligible residents of California. For example, only 69 of the 120 adult interviews completed from the Vietnamese surname list involved respondents who indicated being having Vietnamese ethnicity.

<sup>2</sup> Includes interviews meeting the criteria as partially complete.

<sup>3</sup> Breakdown of completes by frame deviates slightly from original sample numbers due to numbers changing frames following post-sampling database processing.

Interviews in all languages were administered using SSRS's computer-assisted telephone interviewing (CATI) system. The average adult interview took about 43 minutes to complete. The average child and adolescent interviews took about 19 minutes and 24 minutes, respectively. For "child-first" interviews, additional household information asked as part of the child interview averaged about 14 minutes. Interviews in non-English languages typically took longer to complete with an average length of about 53 minutes for the adult interview, 31 minutes for the teen, and 21 minutes for the child. More than seven 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 seven percent of all adolescent interviews.

Table 1-3 shows the major topic areas for each of the three survey instruments (adult, child, and adolescent).

Table 1-3. CHIS 2017 survey topic areas by instrument

<b>Health status</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
General health status	✓	✓	✓
Days missed from school due to health problems		✓	✓
<b>Health conditions</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Asthma	✓	✓	✓
Diabetes, gestational diabetes, pre- /borderline diabetes	✓		
Heart disease, high blood pressure, stroke	✓		
Physical, behavioral, and/or mental conditions			✓
Physical disabilities, blindness, deafness	✓		
<b>Mental health</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Mental health status	✓	✓	
Perceived need, access and utilization of mental health services	✓	✓	
Suicide ideation and attempts	✓	✓	
Functional impairment, stigma	✓		
<b>Health behaviors</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Dietary intake and soda intake	✓	✓	✓
Water consumption	✓	✓	
Physical activity and exercise	✓	✓	✓
Commute from school to home		✓	✓
Sedentary time		✓	✓
Walking for transportation and leisure	✓		
Alcohol and drug use	✓	✓	
Cigarette and E-cigarette use	✓	✓	
Sexual behavior	✓	✓	
Breastfeeding			✓
Sleep and Technology		✓	
Sedentary Time		✓	✓
<b>Women's health</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Pregnancy	✓		
<b>Dental health</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Last dental visit, main reason haven't visited dentist	✓	✓	✓
Current dental insurance coverage	✓		✓

(continued)

Table 1-3. CHIS 2017-2018 survey topic areas by instrument (continued)

<b>Neighborhood and housing</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Safety, social cohesion	✓	✓	✓
Homeownership, length of time at current residence	✓		
Park use		✓	✓
Civic engagement	✓	✓	
Building Healthy Communities	✓		
<b>Access to and use of health care</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Usual source of care, visits to medical doctor	✓	✓	✓
Emergency room visits	✓	✓	✓
Delays in getting care (prescriptions and medical care)	✓	✓	✓
Medical home, timely appointments, hospitalizations	✓	✓	✓
Developmental screening			✓
Communication problems with doctor	✓		✓
Tele-medical care	✓		
Family planning	✓		
Change of usual source of care	✓		
<b>Food environment</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Access to fresh and affordable foods	✓		
Availability of food in household over past 12 months	✓		
Hunger	✓		
School lunch consumption		✓	
<b>Health insurance</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Current insurance coverage, spouse's coverage, who pays for coverage	✓	✓	✓
Health plan enrollment, characteristics and plan assessment	✓	✓	✓
Whether employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months, reasons for lack of insurance	✓	✓	✓
Difficulty finding private health insurance	✓		
High deductible health plans	✓	✓	✓
Partial scope Medi-Cal	✓		
<b>Public program eligibility</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Household poverty level	✓		
Program participation (CalWORKs, Food Stamps, SSI, SSDI, WIC, TANF)	✓	✓	✓
Assets, alimony/child support, social security/pension, worker's compensation	✓		
Reason for Medi-Cal non-participation among potential beneficiaries	✓	✓	✓
Medi-Cal eligibility	✓	✓	✓



Table 1-3. CHIS 2017-2018 survey topic areas by instrument (continued)

<b>Bullying and interpersonal violence</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Bullying, personal safety, school safety, interpersonal violence		✓	
<b>Parental involvement/adult supervision</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Adult presence and support		✓	
Technology use		✓	
<b>Child care and school attendance</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Current child care arrangements			✓
Paid child care	✓		
Preschool/school attendance, name of school		✓	✓
Preschool quality			✓
School instability		✓	
First 5 California: "Talk, Read, Sing Program"			✓
<b>Employment</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Employment status, spouse's employment status	✓		
Hours worked at all jobs	✓		
Industry and occupation, firm size	✓		
<b>Income</b>	<b>Adult</b>	<b>Teen</b>	<b>Child</b>
Respondent's and spouse's earnings last month before taxes	✓		
Household income, number of persons supported by household income	✓		
<b>Respondent characteristics</b>	<b>Adult</b>	<b>Teen</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		✓	
Education, English language proficiency	✓		
Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home	✓	✓	✓
Education of primary caretaker			✓
Citizenship, immigration status, country of birth, and length of time in U.S. of parents			✓

Source: UCLA Center for Health Policy Research, 2017 California Health Interview Survey.

## 1.5 Response Rates

The overall response rates for CHIS 2017 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 2017, the landline/list sample household response rate was 9.3 percent (the product of the screener response rate of 13.2 percent and the extended interview response rate at the household level of 70.3 percent). The cell sample household response rate was 6.5 percent, incorporating a screener response rate of 10.0 percent household-level extended interview response rate of 65.2 percent. CHIS uses AAPOR response rate RR4 (see more detailed in *CHIS 2017 Methodology Series: Report 4 – Response Rates*).

Within the landline and cell phone sampling frames for 2017, the extended interview response rate for the landline/list sample varied across the adult (61.0 percent), child (63.3 percent) and adolescent (26.6 percent) interviews. The adolescent rate includes the process of obtaining permission from a parent or guardian.

The adult interview response rate for the cell sample was 66.6 percent, the child rate was 63.9 percent, and the adolescent rate was 20.3 percent in 2017 (see Table 1-4a). 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 each survey year (see Table 1-4b). As in previous years, household and person level response rates vary by sampling stratum. CHIS response rates are similar to, and sometimes higher than, other

Table 1-4a. CHIS response rates - Conditional

Type of Sample	Screener	Household	Adult (given screened)	Child (given screened & eligibility)	Adolescent (given screened & permission)
Overall	11.0%	66.8%	63.0%	63.7%	23.4%
Landline RDD/List	13.2%	70.3%	61.0%	63.3%	26.6%
Cell RDD/List	10.0%	65.2%	66.6%	63.9%	20.3%

Source: UCLA Center for Health Policy Research, 2017 California Health Interview Survey. comparable surveys that interview by telephone.

Table 1-4b. CHIS response rates - Unconditional

Type of Sample	Screener	Household	Adult (given screened)	Child (given screened & eligibility)	Adolescent (given screened & permission)
Overall	11.0%	7.1%	6.7%	6.7%	2.4%
Landline RDD/List	13.2%	9.3%	8.1%	8.4%	3.5%
Cell RDD/List	10.0%	6.5%	6.6%	6.4%	2.0%

Source: UCLA Center for Health Policy Research, 2017 California Health Interview Survey.

To maximize the response rate, especially at the screener stage, an advance letter in six languages was mailed to all landline sampled telephone numbers for which an address could be obtained from reverse directory services. An advance letter was mailed for 36.5 percent of the landline RDD sample telephone numbers not identified by the sample vendor as business numbers or not identified by SSRS’s dialer software as nonworking numbers, and for 100 percent of surname list sample numbers. Combining these two frames, advance letters were sent to 36.9 percent of all fielded landline telephone numbers. For cell sample, an advance letter was mailed for 38.5 percent of the RDD sample telephone numbers not identified by the sample vendor as business numbers or not identified by SSRS’s dialer software as nonworking numbers, and for 100 percent of surname list sample numbers. Combining these two frames, advance letters were sent to 40.7 percent of all fielded cell telephone numbers. As in all CHIS cycles since CHIS 2005, a \$2 bill was included with the CHIS 2017 advance letter to encourage cooperation. Unlike previous cycles, additional incentives were not offered to cell phone and nonresponse follow up (NRFU) respondents.

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 2017, either a spouse/partner or adult child completed a proxy interview for 3 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 2017 accomplish the following objectives:

- Compensate for differential probabilities of selection for phone numbers (households) and persons within household;
- Reduce biases occurring because nonrespondents may have different characteristics than respondents;
- Adjust, to the extent possible, for undercoverage in the sampling frames 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 telephone number) and a variety of 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 2017 were created primarily from the California Department of Finance’s (DOF) 2017 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 Region in San Diego County), and education. One limitation of using Department of Finance (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 Department of Finance control totals prior to calibration.

The DOF control totals used to create the CHIS 2017 weights are based on 2010 Census counts, as were those used for the 2015-2016 cycle. Please pay close attention when comparing estimates using CHIS 2017 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 nonrespondents 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, SPA, HSR, 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 2017 were low, with most variables missing valid responses for less than 1% of the sample.

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 continuous 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 and region of California were imputed by SSRS. UCLA-CHPR began their imputation process by imputing household income and educational attainment, so that these characteristics are 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.

#### **4. Sample Code for Analysis and Pooling of CHIS Data**

As previously noted, sample code to assist with analyses and pooling of CHIS data is available on the CHIS website at <http://healthpolicy.ucla.edu/chis/analyze/Pages/default.aspx>

## 5. Restricted Variables

The following geographic variables are not located in the funder files, but may be accessible upon request and IRB approval (UCLA & CPHS). These variables are restricted due to their identifiable nature.

VARIABLE	LABEL	NOTE
LATITUDE	LATITUDE	GEOGRAPHIC LONGITUDE OF RESIDENCE. USED FOR MAPPING.
LONGIT	LONGITUDE	GEOGRAPHIC LONGITUDE OF RESIDENCE. USED FOR MAPPING.
CBLK	CENSUS BLOCK INCLUDING CENSUS TRACT	SMALLEST CENSUS DELINEATION AVAILABLE TO MERGE WITH CENSUS DATA; USED FOR MULTILEVEL MODELING, NEIGHBORHOOD ANALYSIS.

### Teen and child restricted variables

VARIABLE	LABEL	NOTE
SCH_BLK	SCHOOL CENSUS TRACT INCLUDING BLOCK	SMALLEST CENSUS DELINEATION AVAILABLE TO MERGE WITH CENSUS DATA; USED FOR MULTILEVEL MODELING, SCHOOL NEIGHBORHOOD ANALYSIS.
SCH_CDS	STATE SCHOOL ID NUMBER	
SCH_LAT	SCHOOL LATITUDE	GEOGRAPHIC LONGITUDE OF SCHOOL. USED FOR MAPPING.
SCH_LON	SCHOOL LONGITUDE	GEOGRAPHIC LONGITUDE OF SCHOOL. USED FOR MAPPING.



## 6. CHIS Data Dictionary

### How to Use the Data Dictionary

This Data Dictionary describes the variables in the CHIS 2017 data file. The index of the data dictionary lists variables first in alphabetical order and then in the order they were administered in the survey questionnaire. Please note that identical variable names appearing in the CHIS 2001, CHIS 2003, CHIS 2005, CHIS 2007, CHIS 2009, CHIS 2011-12, CHIS 2013-2014, and CHIS 2015-16 data files does not guarantee identical question wording, response categories or universe; please consult the questionnaires and data dictionaries to assess comparability across cycles. A printable version of the questionnaire can be found on the California Health Interview Survey web site at <http://healthpolicy.ucla.edu/chis/design/Pages/questionnairesEnglish.aspx>.

The data dictionary contains the following fields:

- **VARNAME:** The names of the variables.
- **QNAME17:** Starting in CHIS 2017, the field is blank. Each questionnaire item will be referred to by its variable name only.
- **QNAME1516:** The names of 2015-2016 survey items identical or similar to the 2017 items.
- **LABEL:** A description (or label) of the variable which is the same as what is included in the label file described in Section 1.2.
- **VALUE:** Response categories and their sample distributions of categorical variables. The following negative values are used for all variables:
  - 1: INAPPLICABLE.
  - 2: PROXY SKIPPED.
  - 5: CHILD/HOUSEHOLD INFORMATION NOT COLLECTED FOR TEEN AND CHILD INTERVIEWS.
  - 7: REFUSED.
  - 8: DON'T KNOW.
  - 9: NOT ASCERTAINED.
- **MEAN STATISTICS:** Sample distributions of continuous variables.
- **UNIVERSE:** The scope of eligible respondents for the corresponding item. For some questions and variables, certain respondents become ineligible due to skip patterns or other restrictions (e.g., age and sex).
- **INPUT VAR:** Source variables used to construct the one in the **VARNAME** field.
- **NOTES:** Additional information about the variable.

### Navigation

Users can also navigate between data dictionary sections by clicking on the section listings on the next page.