



california
health
interview
survey

CHIS 2014 DATA DICTIONARY

ÚWØ Data Dictionary

Teen

U&q à^!ÁGEFì



california
health
interview
survey

CHIS 2014 DATA DICTIONARY

Teen Survey

RWH

Qevqdt'423:

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

Suggested Citation:

California Health Interview Survey. CHIS 2014 Teen Survey. UCLA Center for Health Policy Research. Los Angeles, CA: Qevqdt"423:

TABLE OF CONTENTS

1. Introduction	2
1.1 CHIS 2014 Teen Survey Data File	2
1.2 Accompanying Files	2
2. What’s New in CHIS 2013-14.....	2
3. CHIS 2013-2014 Design and Methodology Summary	5
4. Sample Code for Analysis and Pooling of CHIS data.....	13
5. Restricted Variables	14
6. CHIS Data Dictionary	15

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



1. Introduction

1.1 CHIS 2014 Teen Survey Data File

The 2014 California Health Interview Survey (CHIS 2014) Teen Data File consists of individual records obtained from the 2014 data collection period of the CHIS 2013-2014 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 dacchpr@ucla.edu. Limited technical assistance is also available from CHIS – please send email to 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, teen.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
See Section 3.7 for descriptions of imputation flag variables and values.
- f. **Others:** TEEN.XPT, TEEN.XPT

2. What's New in CHIS 2013-14

As an ongoing cross-sectional survey of California's population, each CHIS 2-year data cycle has its own unique features. **This document describes the main cycle-specific methodological changes that were implemented in CHIS 2013-2014.** It is recommended that CHIS data users review the information below and detailed online documentation as necessary before analyzing or reporting CHIS data.

CHIS Methodological Documentation Online:

<http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx>

- 1) **One-Year Data File Release** – CHIS 2013 and 2014 public use files (PUFs) were released as separate, one-year data files (one file for each year), which is different from the combined two-year data files released in the past. The switch to single-year data files was facilitated by continuous data collection, which began in the CHIS 2011-2012 data collection cycle. The CHIS 2013-2014 data collection cycle includes interviews conducted between February 2013 and early January 2015 with approximately half of the interviews conducted during calendar year 2013 and the other half during calendar year 2014. We will re-release CHIS 2011 and 2012 data files as separate, one-year files later in 2015 for researchers who want to make single year estimates across this four-year timespan. Two-year data files for the CHIS 2013-2014 cycle can be requested later this year via our Data Access Center (DAC, see <http://healthpolicy.ucla.edu/chis/data/Pages/GetCHISData.aspx>).

As of the CHIS 2013 and 2014 data file release (August 2015), AskCHIS (<http://ask.chis.ucla.edu>) included individual year data for CHIS 2011, 2012, 2013, and 2014, so users can produce single-year estimates. In future years, we anticipate releasing one-year PUFs and updating AskCHIS annually, which will provide CHIS users with much more timely data and greater flexibility in structuring their analyses. Users should be cautious examining indicators for small populations (such as child, teen, or racial/ethnic groups) due to the smaller sample sizes of the one-year data; pooling two or more cycles of one-year data is generally advised.

Users who need more information about pooling/trending data over time should review the Analyze CHIS Data website at <http://healthpolicy.ucla.edu/chis/analyze/Pages/default.aspx> or go to the Analyze CHIS Data user forum at <http://healthpolicy.ucla.edu/forum/Pages/Forum.aspx>.

- 2) **Measuring Modified Adjusted Gross Income (MAGI) and Medi-Cal Eligibility** – CHIS has always included adult survey questions to estimate eligibility for California's Medicaid program, Medi-Cal. The 2010 Patient Protection and Affordable Care Act (ACA) included numerous changes to Medi-Cal eligibility. Effective January 1, 2014, income-based eligibility for Medi-Cal and Healthy Families (California's Children's Health Insurance Program) participation is determined by Modified Adjusted Gross Income (MAGI). CHIS 2014 (and CHIS 2013, but most relevant to 2014) included the following changes to approximate a respondent's MAGI.
 - a. Added two questions to identify households receiving workers' compensation in the past month and the amount received in order to exclude workers' compensation income from total income (AL32, AL33);
 - b. Modified questions on child support to eliminate "other" income sources (government or veterans' programs) (AL15, AL16) to exclude child support income from total income;
 - c. Added two questions clarifying whether there is anyone else not living in the household, but living in the U.S., who is supported by the total household income reported (AK32, AK33).

The MAGI approximation informed the construction of the CHIS 2014 variable, **ELGMAGI3**, capturing uninsured individuals who are “newly eligible” for Medi-Cal under the ACA.

ELGMAGI3 also categorizes many uninsured individuals eligible for Healthy Families as *Medi-Cal eligible* due to the transition of Healthy Families enrollees into the Targeted Low Income Medicaid Program.

Detailed documentation about the MAGI variable is forthcoming. See the Analyze CHIS Data forum for more information (<http://healthpolicy.ucla.edu/forum/Pages/Forum.aspx>).

- 3) **New and Updated Survey Questions** – Most CHIS questions are included in every CHIS cycle, but some are added or removed depending on both stake-holder input on public health importance and funding availability. Noteworthy changes include:
 - a. New content CHIS 2013-2014:
 - i. 2014 only: Questions to determine Covered California (the state health insurance marketplace) participation and experience shopping for coverage in the individual market.
 - ii. 2013-2014: Contraception use, access to general and specialty care, dental insurance, industry and occupation, teen and child sedentary behavior, and teen grade level and organizational involvement.
 - b. Removed content: falls among older adults; moderate and vigorous activity in the past week; fruit and vegetable consumption in the past month; mammography; teen drug use/sexual behavior and STI testing; and, bullying and interpersonal violence among teens.
- 4) **Japanese American Oversample and Tagalog Interviewing** – A special oversample of Japanese Americans was conducted using telephone numbers associated with Japanese first and last names. These phone numbers were used to increase the yield of Californians of Japanese descent for an oversample of about 130 Japanese Americans. The *CHIS 2013-2014 Methodology Report 1 – Sample Design* includes more information about this oversample. CHIS continued to oversample Koreans and Vietnamese as well. CHIS 2013-2014 also includes interviews in Tagalog, in addition to its usual languages: English, Spanish, Chinese (Cantonese and Mandarin dialects), Korean and Vietnamese.
- 5) **County Oversamples** – As before, some counties funded additional interviews. In CHIS 2013-2014, approximately 1,600 additional households were sampled in San Diego County to provide sub-county estimates (as has been done since 2005). Three other counties also included supplemental oversamples (Calaveras, Siskiyou, and Tuolumne) of about 400 additional households each. An address-based sample was used in Sonoma County to complete interviews in about 500 additional households.
- 6) **Dual-Frame Cell Phone & Landline RDD Sample** – In CHIS 2013-2014, 7,752 adult interviews were conducted from the cell phone sample (19.3% of adult interviews). In CHIS 2011-2012, 9,151 adult interviews were conducted from the cell phone sample (21.3% of adult interviews). CHIS 2009 had 3,047 (6.4% of adult interviews) interviews conducted from the cell phone sample. Like CHIS 2011-2012, CHIS 2013-2014 used county-level goals for the cell phone RDD sample. The complete sample design is in *CHIS 2013-2014 Methodology Report 5 – Sample Design* (to be released Fall 2015). For more about cell phone sampling from the American Association for Public Opinion Research see http://www.aapor.org/Cell_Phone_Task_Force_Report.htm

3. CHIS 2013-2014 Design and Methodology Summary

3.1 Overview

A series of five methodology reports will be available with more detail about the methods used in CHIS 2013-2014:

- 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.

The CHIS is a population-based telephone survey of California's population conducted every other year since 2001 and continually beginning in 2011. CHIS is the largest state health survey and one of the largest health surveys in the nation. CHIS is conducted by the UCLA Center for Health Policy Research (UCLA-CHPR) in collaboration with the California Department of Public Health and the 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 to meet and optimize 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).

3.2 Switch to a Continuous Survey

From the first CHIS cycle in 2001 through 2009, CHIS data was collected during a 7-9 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 2013-2014 data were collected between February 2013 and early January 2015. Approximately half of the interviews were conducted during the 2013 calendar year and half during the 2014 calendar year. 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 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 code, is available at: <http://healthpolicy.ucla.edu/chis/analyze/Pages/sample-code.aspx>.

Additional documentation on constructing the CHIS sampling weights is available in CHIS Methods Report #5—Weighting and Variance Estimation, available at: <http://healthpolicy.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.

3.3 Sample Design Objectives

The CHIS 2013-2014 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 random-digit-dial (RDD) sample was designed to achieve completed adult interviews with approximately 80% landline and 20% cellular phone numbers. For the landline 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 Los Angeles County stratum included 8 sub-strata for Service Planning Areas, and the San Diego County stratum included 6 sub-strata for Health Service Districts. Approximately 1,600 additional households were sampled in San Diego County to obtain the sub-county estimates. Most of the strata (39 of 44) are made up of a single county with no sub-strata (counties 3-41 in Table 1-1), with three multi-county strata comprised of the 17 remaining counties (see Table 1-1). CHIS 2013-2014 also included supplemental geographic oversamples of landlines in 3 small counties (Calaveras, Siskiyou, and Tuolumne) that were part of multi-county strata (about 400 additional households for each county). An address-based sample of an additional 500 households was conducted in Sonoma County and oversamples of about 130 Japanese Americans, 104 Korean Americans, and 120 Vietnamese Americans were completed using list samples. A sufficient number of adult interviews were allocated to each stratum and sub-stratum to support the first sample design objective—to provide health estimates for adults at the local level. The same landline geographic stratification of the state has been used since CHIS 2005.

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 were randomly selected; the adolescent was interviewed directly, and the adult most knowledgeable about the child's health completed the child interview.

The RDD CHIS 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 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 further increase the sample size for Koreans and Vietnamese. Surname and given name lists were used similarly to increase the yield of Californians of Japanese descent.

Table 3-1. California county and county group strata used in the CHIS 2013-2014 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, Glen, Tehama
3. Orange	23. San Luis Obispo	43. Plumas, Sierra, Siskiyou, Lassen, Modoc, Trinity, Del Norte
4. Santa Clara	24. Placer	44. Mariposa, Mono, Tuolumne, Alpine, Amador, Calaveras, Inyo
5. San Bernardino	25. Merced	
6. Riverside	26. Butte	

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

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 2013-2014, the goal was to complete approximately 8,000 interviews (20% of all RDD interviews statewide) with adults from the cell phone sample. Although the geographic information available for cell phone numbers is not as precise as that for landlines, cell phone numbers were assigned to the same 44 strata (i.e., 41 strata defined by a single county and 3 strata created by multiple counties). 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. As in CHIS 2011-2012, if a sampled cell number was shared by two or more adult members of a household, one household member was selected for the adult interview; otherwise the adult owner of the sampled number was selected. Cell numbers used exclusively by children under 18 were considered ineligible. About 480 teen interviews and 1,250 child interviews were completed from the cell phone sample in CHIS 2013-2014.

The cell phone sampling method used in CHIS has evolved since its first implementation in 2007 when only cell numbers belonging to adults in cell-only households were eligible for sampling adults. There have been two 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 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 7 geographic sampling strata created using telephone area codes. In CHIS 2009 and 2011-2012, the number of cell phone strata was increased to 28. These strata were created using both area codes and the geographic information assigned to the number. In CHIS 2011-2013, 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.

3.4 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, for the first time, 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.

Westat, a private firm that specializes in statistical research and large-scale sample surveys, conducted CHIS 2013-2014 data collection under contract with the UCLA Center for Health Policy Research. For all samples, Westat 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. In landline, list, and ABS sample households with children where 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 2013-2014 by the type of sample (landline RDD, surname list, cell RDD, and Sonoma ABS). These numbers are provided in greater detail in *CHIS 2013-2014 Methodology Series: Report 2 – Data Collection*. Note that these figures were accurate as of data collection completion 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://healthpolicy.ucla.edu/chis/design/Pages/sample.aspx>.

Table 3-2. Number of completed CHIS 2013-2014 interviews by type of sample and instrument

Type of sample	Adult*	Child	Adolescent
Total all samples	40,240 ¹	5,512	2,253
Landline RDD	31,615	4,164	1,738
Surname list	392	50	18
Cell RDD	7,752	1,256	482
Sonoma ABS	481	42	15

*Includes interviews meeting the criteria as partially complete

Source: UCLA Center for Health Policy Research, 2013-2014 California Health Interview Survey.

Interviews in all languages were administered using Westat's computer-assisted telephone interviewing (CATI) system. The average adult interview took about 36 minutes to complete. The average child and adolescent interviews took about 16 minutes and 23 minutes, respectively. For "child-first" interviews, additional household information asked as part of the child interview averaged about 9 minutes. Interviews in non-English languages generally took longer to complete. More than 11 percent of the adult interviews were completed in a language other than English, as were about 23 percent of all child (parent proxy) interviews and 5 percent of all adolescent interviews.

3.5 Response Rates

The overall response rate for CHIS 2013-2014 is a composite 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). To maximize the response rate, especially at the screener stage, an advance letter in five 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 50.7 percent of the landline RDD sample telephone numbers not identified by the sample vendor as business or nonworking numbers, and for 82.2 percent of surname list sample numbers. Addresses were not available for the cell sample. As in all CHIS cycles since CHIS 2005, a \$2 bill was included with the CHIS 2013-2014 advance letter to encourage cooperation.

¹Numbers in this table represent the data publically released and available through our Data Access Center. Total sample sizes may differ for specific calculations within the five methodology reports, or for specific analyses based on CHIS data.

The CHIS 2013-2014 screener response rate for the landline/list sample was 28.8 percent, and was higher for households that were sent the advance letter. For the cell phone sample, the screener response rate was 30.7 percent. The extended interview response rate for the landline/list sample varied across the adult (44.8 percent), child (68.9 percent) and adolescent (40.2 percent) interviews. The adolescent rate includes getting permission from a parent or guardian. The adult interview response rate for the cell sample was 52.1 percent, the child rate was 72.2 percent, and the adolescent rate 41.0 percent. Multiplying the screener and extended rates gives an overall response rate for each type of interview. The percentage of households completing one or more of the extended interviews (adult, child, and/or adolescent) is a useful summary of the overall performance of the landline sample. For CHIS 2013-2014, the landline/list sample household response rate was 14.8 percent (the product of the screener response rate and the extended interview response rate at the household level of 51.4 percent). The cell sample household response rate was 16.6 percent, incorporating a household-level extended interview response rate of 53.9 percent. All of the household and person level response rates vary by sampling stratum.

Historically, the CHIS response rates are comparable to response rates of other scientific telephone surveys in California, such as the California Behavioral Risk Factor Surveillance System (BRFSS) Survey. However, comparing the CHIS and BRFSS response rates requires recomputing the CHIS response rates so they match the BRFSS response rate calculation methods. The 2013 California BRFSS landline response rate is 38.9 percent, the cell phone response rate is 39.3 percent, and the combined landline and cell phone rate is 39.0 percent.² Recalculating the CHIS response rates using the BRFSS method, the CHIS 2013-2014 landline response rate is 39.5, cell phone response rate is 32.1 percent, and the combined landline and cell phone response rate is 37.2 percent. California as a whole and the state's urban areas in particular are among the most difficult parts of the nation in which to conduct telephone interviews. For example, based on the last reported BRFSS refusal rates in 2011; the refusal rate for California (31.4%) was the highest in the nation and was twice the national median (16.0%). Survey response rates tend to be lower in California than nationally, and over the past decade response rates have been declining both nationally and in California. For more information about the CHIS 2013-2014 response rates please see *CHIS 2013-2014 Methodology Series: Report 4 – Response Rates*.

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

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 (see section 2.6 on imputation methods for more information).

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. For 248 adults, a proxy interview was completed by either a spouse/partner or adult child. A reduced questionnaire, with questions identified as appropriate for a proxy respondent, was administered.

3.6 Weighting the Sample

² As reported in the Behavioral Risk Factor Surveillance System: 2013 Summary Data Quality Report. Retrieved May 22, 2015, available online at http://www.cdc.gov/brfss/annual_data/2013/pdf/2013_dqr.pdf

To produce population estimates from CHIS data, weights are 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 is weighted to represent the non-institutionalized population for each sampling stratum and statewide. The weighting procedures used for CHIS 2013-2014 accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons;
- Reduce biases occurring because non-respondents may have different characteristics than respondents;
- Adjust, to the extent possible, for under-coverage 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 is used to compute a person-level weight, which includes adjustments for the within-household sampling of persons and nonresponse. The final step is to adjust the person-level weight using an iterative proportional fitting method, or “raking” as it is commonly called, so that CHIS estimates are consistent with the marginal population control totals. This iterative procedure forces the CHIS weights to sum to known population control totals from an independent data source (see below). The procedure requires iteration to make sure all the control totals, or raking dimensions, are simultaneously satisfied within a pre-specified tolerance.

Population control totals of the number of persons by age, race, and sex at the stratum level for CHIS 2013-2014 were created primarily from the California Department of Finance’s (DOF) 2014 Population Estimates and 2014 Population Projections. The raking procedure used 12 raking 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), household composition (presence of children and adolescents in the household), and socio-economic variables (home ownership and education). The socio-economic variables are included to reduce biases associated with excluding households without landline telephones from the sample frame. 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 raking.

The 2014 DOF control totals used to create the CHIS 2013-2014 weights are based on 2010 Census counts, as were those used for the 2011-2012 cycle. Please pay close attention when comparing estimates using CHIS 2013-2014 data with estimates using data from CHIS cycles before 2010. The most accurate California population figures are available when the US population count is conducted (every 10 years). Population-based surveys like CHIS must use estimates and projections based on the decennial population count data between Censuses. 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. See *CHIS 2013-2014 Methodology Series: Report 5 – Weighting and Variance Estimation* for more information on the weighting process.

3.7 Imputation Methods

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

Two different imputation procedures were used by Westat 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 without replacement. The hot deck approach is one of the most commonly used methods for assigning values for missing responses. With a hot deck, a value reported by a respondent for a particular item is 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 answer a survey item form a pool of donors, while the item non-respondents form a group of recipients. A recipient is matched to the subset pool of donors based on household and individual characteristics. A value for the recipient is then randomly imputed from one of the donors in the pool. Once a donor is used, it is removed from the pool of donors for that variable. Westat used hot deck imputation to impute the same items in all CHIS cycles since 2003 (i.e., race, ethnicity, home ownership, and education).

UCLA-CHPR imputed missing values for nearly every variable in the data files other than those imputed by Westat and some sensitive variables in which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2013-2014 were low, with most variables missing valid responses for less than 2% of the sample. However, there were a few exceptions where item nonresponse rate was greater than 20%, such as household income.

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 replaces 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 corresponds to the nature of the variable being imputed (e.g. linear regression for continuous variables, logistic regression for binary variables, etc.). Donors and recipients are 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 Westat. UCLA-CHPR begins 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 is 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 & CHPS). 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

This Data Dictionary describes the variables in the CHIS 2014 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, and CHIS 2013-14 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/questionnaires.aspx>.

The data dictionary contains the following fields:

- **VARNAME:** The names of the variables.
- **QNAME14:** The names of the 2014 survey items. A blank/NA field means the variable is constructed with survey items shown in the **INPUT VAR** field.
- **QNAME13:** The names of the 2013 survey items identical or similar to the 2013-14 items. A blank/NA field means the variable is constructed with survey items shown in the **INPUT VAR** field.
- **QNAME11:** The names of 2011-12 survey items identical or similar to the 2013-14 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.