



December 13, 2016

CHIS 2015 Methodology Report Series

Report 3

Data Processing

-Short Report-

Table of Contents: CHIS 2015 Methodology Short Report #3

| | |
|---|----|
| Chapter 1: CHIS 2015 Methodology Overview | 3 |
| 1.1 Overview | 3 |
| 1.2 Switch to a Continuous Survey | 4 |
| 1.3 Sample Design Objectives | 4 |
| 1.4 Data Collection..... | 7 |
| 1.5 Response Rates..... | 8 |
| 1.6 Weighting the Sample | 9 |
| 1.7 Imputation Methods | 11 |
| Chapter 2: Detailed Results from Data Processing..... | 13 |
| CHIS 2015 Sampling Strata | 13 |
| CHIS 2015 Completed Interviews by sampling frame and instrument | 14 |
| CHIS 2015 Survey Topics..... | 15 |

Chapter 1: CHIS 2015 Methodology Overview

CHIS 2015 SAMPLE DESIGN AND METHODOLOGY SUMMARY

-SHORT REPORT-

1.1 Overview

CHIS has historically released 5 methodology reports with each cycle's data release. With the move to annual data release we are releasing reduced versions of those reports following the same structure listed below. This documentation covers the first half (CHIS 2015) of the CHIS 2015-2016 cycle.

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

Each of these short reports begins with the same summary chapter, which includes highlights of various methodological components of the survey. The rest of each chapter includes additional documentation on that aspect of the methodology. The full series of complete methodology reports will be available in 2017 with more detail about the methods used in CHIS 2015-2016.

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. For methodology reports from previous CHIS cycles, go to <http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx>

CHIS is a population-based telephone survey of California's residential, non-institutionalized 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. 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 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 7-9 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 2015 data were collected between May 2015 and mid-February 2016. At the writing of this document and release of CHIS 2015 data, CHIS 2016 is still in the field. 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 2015-2016, data users will be able to produce single-year estimates using the weights provided (referred to as CHIS 2015 and CHIS 2016, respectively). This is a new feature of CHIS data.

See what's new in the 2015-2016 CHIS sampling and data collection here:
<http://healthpolicy.ucla.edu/chis/design/Documents/whats-new-chis-2015.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://healthpolicy.ucla.edu/chis/analyze/Pages/sample-code.aspx>.

Additional documentation on constructing the CHIS sampling weights is available in the forthcoming CHIS 2015-2016 Methods Report #5—Weighting and Variance Estimation, which will be posted at <http://healthpolicy.ucla.edu/chis/design/Pages/methodology.aspx> once available. The 2015 short report provides initial information on weight construction (available at the same URL). 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 2015-2016 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 completed adult interviews via approximately 50% landline and 50% cellular phone numbers. 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 8 sub-strata for Service Planning Areas, and the San Diego County stratum included 6 sub-strata for Health Service Districts. Most of the strata (39 of 44) consisted 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). An additional sample from both the landline and cell phone frames produced 1,042 interviews within Marin County. An Asian surname sample list frame households also produced additional respondents: 173 Japanese, 146 Korean, and 234 Vietnamese adult interviews based on self-identified ethnicity. Overall, 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.

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 most knowledgeable about the child's health completed the child interview.

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 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 1-1. California county and county group strata used in the CHIS 201516 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, |

| | | |
|-------------------|------------|-----------------------------------|
| 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, 2015-2016 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 2015, the goal was to complete approximately 10,222 interviews (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. An adult reached on a sampled non-business cell phone number was automatically selected for CHIS (i.e., no within-household sampling for the adult interview, but child and teen interviews were possible using the same relationship rules as the landline sample). Cell numbers used exclusively by children under 18 were considered ineligible. A total of 754 teen interviews and 2,157 child interviews were completed in CHIS 2015 with approximately 46% 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.¹ 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 interviews. 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.

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 used in the cell phone sample 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. 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 2015.

¹ <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201605.pdf>

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 is the first year that Tagalog interviews have been conducted from the beginning. 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.

RTI International designed the methodology and collected data for CHIS 2015, under contract with the UCLA Center for Health Policy Research. They are an independent, nonprofit institute that provides research, development, and technical services to government and commercial clients worldwide, with specialization in designing and implementing large-scale sample surveys. For all sampled households, RTI staff interviewed one randomly selected adult, 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 retained 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 some 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 2015 by the type of sample (landline RDD, surname list, cell RDD, and ABS). 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 1-2. Number of completed CHIS 2015 interviews by type of sample and instrument

| Type of sample ¹ | Adult ² | Child | Adolescent |
|--------------------------------------|--------------------|-------|------------|
| Total all samples | 21,034 | 2,157 | 754 |
| Landline RDD | 7,236 | 660 | 240 |
| Vietnamese surname list | 3,395 | 301 | 105 |
| Korean surname list | 311 | 22 | 10 |
| Japanese surname list | 28 | 2 | 3 |
| Cell RDD | 9,022 | 1,089 | 363 |
| Marin County Oversample ³ | 1,042 | 83 | 33 |

¹ 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. Interviews from the Marin County oversample include respondents who did not live in this county and interviews from the Vietnamese, Korean, or Japanese surname lists include respondents who do not have one of these ethnicities. For example, only 234 of the 3,395 adult interviews completed from the Vietnamese surname list involved respondents who indicated being having Vietnamese ethnicity.

² Includes interviews meeting the criteria as partially complete.

³ Completed interviews for the Marin County oversample do not include interviews completed via the Vietnamese surname list frame. These interviews are counted in the row for the Vietnamese surname list.

Source: UCLA Center for Health Policy Research, 2015-2016 California Health Interview Survey.

Interviews in all languages were administered using RTI'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.

1.5 Response Rates

The overall response rate for CHIS 2015 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). For CHIS 2015, the landline/list sample household response rate was 12.3 percent (the product of the screener response rate of 28.4 and the extended interview response rate at the household level of 43.2 percent). The cell sample household response rate was 9.5 percent, incorporating a screener response rate of 20.7 percent household-level extended interview response rate of 45.9 percent. CHIS uses AAPOR response rate RR4 (see more detailed in *Methodology Report #4 – Response Rates*).

Looking within landline and cell phone sampling frames, the extended interview response rate for the landline/list sample varied across the adult (41.8 percent), child (45.3 percent) and adolescent (17.0 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 48.5 percent, the child rate was 43.4 percent, and the adolescent rate 16.7 percent (see Table 1-3a). Multiplying these rates by the screener response rates used in the household rates above gives an overall response rate for each type of interview (see Table 1-3b). 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 comparable surveys that interview by telephone.

Table 1-3a. CHIS 2015 Response Rates – Conditional

| Type of sample | Screener | Household | Adult (given screened) | Child (given screened) | Adolescent (given screened & permission) |
|----------------|----------|-----------|------------------------------|------------------------------|---|
| Overall | 22.1% | 45.2% | 47.2% | 43.6% | 16.7% |
| Landline RDD | 28.4% | 43.2% | 41.8% | 45.3% | 17.0% |
| Cell RDD | 20.7% | 45.9% | 48.5% | 43.4% | 16.7% |

Table 1-3b. CHIS 2015 Response Rates – Unconditional

| Type of sample | Screener | Household | Adult (given screened) | Child (given screened) | Adolescent (given screened & permission) |
|----------------|----------|-----------|------------------------------|------------------------------|---|
| Overall | 22.1% | 10.0% | 10.4% | 9.6% | 3.7% |
| Landline RDD | 28.4% | 12.3% | 11.9% | 12.9% | 4.8% |
| Cell RDD | 20.7% | 9.5% | 10.0% | 9.0% | 3.4% |

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 20156 advance letter to encourage cooperation. Additional incentives were offered to cell phone and Phase 2 non-response follow up (NRFU) respondents. Details on the incentives can be found in Table 1-4.

Table 1-4. CHIS 2015 Incentives/remuneration by Interview Type

| Type of interview | Amount |
|--|--------|
| <i>Pre-paid</i> | |
| Landline sample matched to address | \$2 |
| <i>Promised</i> | |
| Cell Phone Screener | \$5 |
| Cell Phone Adult Interview | \$20 |
| Cell Phone Child Interview | \$10 |
| Cell Phone Teen Interview | \$10 |
| Non-Response Follow-Up Adult Interview | \$40 |
| Non-Response Follow-Up Child Interview | \$20 |
| Non-Response Follow-Up Teen Interview | \$20 |

We will present a comparison of CHIS 2015-2016 response rates with California BRFSS response rates in the full-cycle 2015-2016 reports. 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.

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 135 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.

1.6 Weighting the Sample

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 2015- accomplish the following objectives:

- Compensate for differential probabilities of selection for phone numbers (households) and persons within household;
- Reduce biases occurring because non-respondents may have different characteristics than respondents;
- Adjust, to the extent possible, for under-coverage in the sampling frames and in the conduct of the survey; and
- Reduce the variance of the estimates by using auxiliary information.
- Account for the second-phase sampling that was part of the responsive and adaptive design (Phase 2 NRFU).

Past CHIS cycles have used a weighting class approach to develop analysis weights. CHIS 2015 uses a model-based approach designed by RTI International. Despite this change in approach, the adjustment dimensions and steps in CHIS 2015 weight development paralleled past cycle approaches as much as possible.

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 for nonresponse. The final step is to adjust the person-level weight using weight calibration, a procedure that forces the CHIS weights to sum to known 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 2015 were created primarily from the California Department of Finance’s (DOF) 2015 Population Estimates and 2015 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 raking.

The 2015 DOF control totals used to create the CHIS 2015 weights are based on 2010 Census counts, as were those used for the 2013-2014 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 U.S. Census Bureau conducts the decennial census. For period 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 massive task designed to enhance the analytic utility of the files. RTI 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 RTI 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. RTI 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 RTI and some sensitive variables in which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2015 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 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 RTI. 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.

Chapter 2: Detailed Results from Data Processing

CHIS 2015 Sampling Strata

Table 2-1 shows the county, sub-county and small county group strata used in CHIS 2015. Similar to past years, CHIS used 46 primary geographic, with eight of those contained within Los Angeles County (strata 1.1-1.8), six contained within San Diego County (strata 2.1-2.6), 39 representing individual counties (strata 3-41), and three being made up of the 14 smallest counties in California (strata 42-44). These strata indicators are used throughout data collection and processing, and for geographic identification at the county level. This table parallels Table 1-1 in CHIS 2013-2014 Methodology Report #3 Data Processing Procedures.

Table 2-1. California county and county group strata used in the CHIS 2015-2016 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, 2015-2016 California Health Interview Survey.

Note: Corresponds to Table 1-1 in CHIS 2013-2014 report.

CHIS 2015 Completed Interviews by sampling frame and instrument

Table 2-2 below shows the number of completed adult, child, and adolescent interviews by sampling frame. Among the completed adult interviews in CHIS 2015, interviews from the cell phone RDD frame represented the largest single source of completed interviews. Adult Interviews completed from landline frames mostly came from the landline RDD frame, but a significant number of adult interviews came from the three Asian ethnicity surname lists - Vietnamese, Korean, and Japanese. The ratio of completed child interviews to adult interviews was somewhat higher for the cell phone RDD frame (0.12) compared to the 4 landline frames, which were all between 0.07 and 0.09. Please note that counts reported here may differ slightly from other reports or from the released data due to the order in which reports were created, changes during data cleaning, etc. This table parallels Table 1-2 in CHIS 2013-2014 Methodology Report #3 Data Processing.

Table 2-2. Number of completed CHIS 2015 interviews by type of sample and instrument

| Type of Sample ¹ | Adult ² | Child | Adolescent |
|--------------------------------------|--------------------|-------|------------|
| Total all samples | 21,034 | 2,157 | 754 |
| Landline RDD | 7,236 | 660 | 240 |
| Cell RDD | 9,022 | 1,089 | 363 |
| Vietnamese surname list | 3,395 | 301 | 105 |
| Korean surname list | 311 | 22 | 10 |
| Japanese surname list | 28 | 2 | 3 |
| Marin County oversample ³ | 1,042 | 83 | 33 |

¹ 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. Interviews from the Marin County oversample include respondents who did not live in this county and interviews from the Vietnamese, Korean, or Japanese surname lists include respondents who do not have one of these ethnicities. For example, only 234 of the 3,395 adult interviews completed from the Vietnamese surname list involved respondents who indicated being having Vietnamese ethnicity.

² Includes interviews meeting the criteria as partially complete.

³ Completed interviews for the Marin County oversample do not include interviews completed via the Vietnamese surname list frame. These interviews are counted in the row for the Vietnamese surname list.

Note: Corresponds to Table 1-2 in CHIS 2013-2014 report.

CHIS 2015 Survey Topics

Table 3 below displays the topics that were asked in CHIS 2015 and their respective instrument/age. This table corresponds to Table 1-3 in CHIS 2013-2014 methodology reports.

In each CHIS cycle we retained about 80% of the content (i.e., survey questions) from previous cycles. For more detail on the specifics of question wording, and changes over cycles, please use the links below.

Full questionnaires (English):

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

Full questionnaires (Other languages):

[http://healthpolicy.ucla.edu/chis/design/Pages/Questionnaires%20\(Translated\).aspx](http://healthpolicy.ucla.edu/chis/design/Pages/Questionnaires%20(Translated).aspx)

Changes in content over cycles: <http://healthpolicy.ucla.edu/chis/analyze/Pages/CHIS-Data-Documentation.aspx>

More details about recoding of specific variable and constructs: Contact our Data Access Center at dacchpr@em.ucla.edu

Table 2-3. CHIS 2015 Survey Topic Areas by Instrument

| | Adult | Teen | Child |
|--|-------------------------------------|-------------------------------------|-------------------------------------|
| Health status | | | |
| General health status | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Days missed from school due to health problems | | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Health-related quality of life (HRQOL) | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| Health conditions | Adult | Teen | Child |
| Asthma | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Diabetes, gestational diabetes, pre- /borderline diabetes | <input checked="" type="checkbox"/> | | |
| Heart disease, high blood pressure, stroke | <input checked="" type="checkbox"/> | | |
| Physical, behavioral, and/or mental conditions | | | <input checked="" type="checkbox"/> |
| Physical disabilities, blindness, deafness | <input checked="" type="checkbox"/> | | |
| Mental health | Adult | Teen | Child |
| Mental health status | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| Perceived need, access and utilization of mental health services | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| Suicide ideation and attempts | <input checked="" type="checkbox"/> | | |

| | | | |
|---|-------------------------------------|-------------------------------------|-------------------------------------|
| Functional impairment, stigma | <input checked="" type="checkbox"/> | | |
| Health behaviors | Adult | Teen | Child |
| Dietary intake, fast food and soda intake | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Water Consumption | | <input checked="" type="checkbox"/> | |
| Physical activity and exercise, commute from school to home | | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Sedentary time | | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Walking for transportation and leisure | <input checked="" type="checkbox"/> | | |
| Doctor discussed nutrition/physical activity | | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Flu Shot | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Alcohol use | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| Cigarette and E-cigarette use | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| Sexual behavior | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| Breastfeeding | | | <input checked="" type="checkbox"/> |
| Women's health | Adult | Teen | Child |
| Mammography screening | <input checked="" type="checkbox"/> | | |
| Pregnancy | <input checked="" type="checkbox"/> | | |
| Dental health | Adult | Teen | Child |
| Last dental visit, main reason haven't visited dentist | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Neighborhood and housing | Adult | Teen | Child |
| Safety, social cohesion | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Homeownership, length of time at current residence | <input checked="" type="checkbox"/> | | |
| Park use | | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Civic engagement | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| Building Healthy Communities | <input checked="" type="checkbox"/> | | |
| Access to and use of health care | Adult | Teen | Child |
| Usual source of care, visits to medical doctor | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Emergency room visits | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Delays in getting care (prescriptions and medical care) | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Medical home, timely appointments, hospitalizations | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Developmental screening | | | <input checked="" type="checkbox"/> |
| Communication problems with doctor | <input checked="" type="checkbox"/> | | <input checked="" type="checkbox"/> |
| Internet use for health information | <input checked="" type="checkbox"/> | | <input checked="" type="checkbox"/> |
| Tele-medical care | <input checked="" type="checkbox"/> | | |
| Family planning | <input checked="" type="checkbox"/> | | |
| Change of usual source of care | <input checked="" type="checkbox"/> | | |

| | Adult | Teen | Child |
|---|-------------------------------------|-------------------------------------|-------------------------------------|
| Food environment | | | |
| Access to fresh and affordable foods | <input checked="" type="checkbox"/> | | |
| Where teen/child eats breakfast/lunch, fast food at school | | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Availability of food in household over past 12 months | <input checked="" type="checkbox"/> | | |
| Hunger | <input checked="" type="checkbox"/> | | |
| Health insurance | | | |
| Current insurance coverage, spouse's coverage, who pays for coverage | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Health plan enrollment, characteristics and plan assessment | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Whether employer offers coverage, respondent/spouse eligibility | <input checked="" type="checkbox"/> | | |
| Coverage over past 12 months, reasons for lack of insurance | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Difficulty finding private health insurance | <input checked="" type="checkbox"/> | | |
| High deductible health plans | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Partial scope Medi-Cal | <input checked="" type="checkbox"/> | | |
| Public program eligibility | | | |
| Household poverty level | <input checked="" type="checkbox"/> | | |
| Program participation (CalWORKs, Food Stamps, SSI, SSDI, WIC, TANF) | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Assets, alimony/child support, social security/pension, worker's compensation | <input checked="" type="checkbox"/> | | |
| Medi-Cal and Healthy Families eligibility | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Reason for Medi-Cal non-participation among potential beneficiaries | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Bullying and interpersonal violence | | | |
| Bullying, personal safety, school safety, interpersonal violence | | <input checked="" type="checkbox"/> | |
| Parental involvement/adult supervision | | | |
| Adult presence after school, role models, resiliency | | <input checked="" type="checkbox"/> | |
| Parental involvement | | <input checked="" type="checkbox"/> | |
| Child care and school attendance | | | |
| Current child care arrangements | | | <input checked="" type="checkbox"/> |
| Paid child care | <input checked="" type="checkbox"/> | | |
| Preschool/school attendance, name of school | | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Preschool quality | | | <input checked="" type="checkbox"/> |
| School instability | | <input checked="" type="checkbox"/> | |
| First 5 California: "Talk, Read, Sing Program" | | | <input checked="" type="checkbox"/> |
| Employment | | | |
| Employment status, spouse's employment status | <input checked="" type="checkbox"/> | | |
| Hours worked at all jobs | <input checked="" type="checkbox"/> | | |

| Income | Adult | Teen | Child |
|--|-------------------------------------|-------------------------------------|-------------------------------------|
| Respondent's and spouse's earnings last month before taxes | <input checked="" type="checkbox"/> | | |
| Household income , number of persons supported by household income | <input checked="" type="checkbox"/> | | |
| Respondent characteristics | Adult | Teen | Child |
| Race and ethnicity, age, gender, height, weight | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Veteran status | <input checked="" type="checkbox"/> | | |
| Marital status, registered domestic partner status (same-sex couples) | <input checked="" type="checkbox"/> | | |
| Sexual orientation | <input checked="" type="checkbox"/> | | |
| Education, English language proficiency | <input checked="" type="checkbox"/> | | |
| Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Education of primary caretaker | | | <input checked="" type="checkbox"/> |
| Citizenship, immigration status, country of birth, and length of time in U.S. of parents | | | <input checked="" type="checkbox"/> |

Note: Corresponds to Table 1-3 in CHIS 2013-2014 report.