



December 13, 2016

CHIS 2015 Methodology Report Series

Report 1

Sample Design

-Short Report-

Table of Contents – CHIS Sample Design Short Methodology Report

Chapter 1: CHIS 2015 Methodology Overview	4
Chapter 2: CHIS 2015-2016 Sampling Design Plan.....	15
Sampling Plan.....	15
Sampling Plan.....	16
1. Background	17
2. Estimation Goals for CHIS 2015-2016	18
2.1 Target Population.....	18
2.2 Analytic Objectives.....	19
3. Sampling Frames	21
3.1 Landline Frame.....	22
3.2 Cell Phone Frame	23
3.3 Supplemental Surname Lists.....	25
4. Household Sampling Design.....	26
4.1 Phase-1 Sample Design and Selection	27
4.1.1 Landline Sample	27
4.1.2 Cell Phone Sample.....	30
4.1.3 Oversampling for Racial Subgroups	32
4.1.4 Supplemental Samples.....	33
4.2 Phase 2 Sample Selection	34
5. Within-Household Sampling	35
5.1 Adult Sampling and Eligibility.....	35
5.2 Selection of Child and Adolescent from Legal Guardian.....	36
5.2.1 Child Sampling.....	37
5.2.2 Adolescent Sampling.....	37
6. Plans for Sample Release	38
7. First Quarter Evaluations And ongoing reports	40
7.1 Study Rates	40
7.2 Sample Vendors	40
7.3 Oversampling for Racial Subgroups	40
7.4 Supplemental Frame Information.....	40
7.5 Quarterly Reports	40
7.6 Weekly Sample Management Reports	41

8. Base Weights.....	42
References	43
Chapter 3: Additional Sampling Documentation	44
Sampled phone numbers by sampling frame	44
Sampled phone numbers by sampling frame and stratum	45
Completed adult interviews by sampling frame and self-reported stratum	47
Completed child interviews by sampling frame and self-reported stratum.....	49
Completed teen interviews by sampling frame and self-reported stratum	51

Figures

Number	Page
Figure 3-1. Transformation of Telephone Frames to Household Population.....	22

Tables

Number	Page
Table 2-1. Design Strata: CHIS 2015-2016	18
Table 2-2. Goals for the Minimum Number of Completed Interviews by Sample Characteristics to Meet Analytic Objectives: CHIS 2015-2016	20
Table 3-1. M-S-G Landline Frame Counts and Estimated Residential Rates (January 2015) by Stratum: CHIS 2015-2016.....	23
Table 3-2. M-S-G Cell Phone Frame Counts by Stratum: CHIS 2015-2016.....	24
Table 4-1. Projected Sample Purchased and Released, and Expected Number of Completed Interviews by Frame, Interview Type, and Design Phase: CHIS 2015-2016	27
Table 4-2. M-S-G Landline Frame Counts (January 2015), Expected Total Sample Size, and Desired Number of Completed Adult Interviews by Stratum: CHIS 2015-2016	28
Table 4-3. M-S-G Cell Phone Frame Counts (January 2015), Expected Total Sample Size, and Desired Number of Completed Adult Interviews by Stratum: CHIS 2015-2016	31
Table 4-4. Base and Supplemental Sample Completed Interviews Targets for Marin County, California: CHIS 2015-2016.....	33

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: CHIS 2015-2016 Sampling Design Plan

Chapter 2 includes the original sampling plan for CHIS 2015-2016. Pages have been re-numbered to conform to this report, but tables and figures have not been renumbered. Note that although the data collection and weighting were designed to produce estimates for 2015 and 2016, the sample design was created for the full two-year cycle.

March 2015

California Health Interview Survey 2015-2016 Sampling Plan

Prepared for

David Grant

Director, California Health Interview Survey (CHIS)

Royce Park

Assistant Director, California Health Interview Survey (CHIS)

UCLA Center for Health Policy Research
10960 Wilshire Blvd, Suite 1550
Los Angeles, CA 90024

Prepared by

Jill Dever

Michael Jacobsen

RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

RTI Project Number 0214657.002.001

RTI Project Number
0214657.002.001

California Health Interview Survey 2015-2016 **Sampling Plan**

March 27, 2015

Prepared for

David Grant

Director, California Health Interview Survey (CHIS)

Royce Park

Assistant Director, California Health Interview Survey (CHIS)

UCLA Center for Health Policy Research
10960 Wilshire Blvd, Suite 1550
Los Angeles, CA 90024

Prepared by

Jill Dever

Michael Jacobsen

RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

1. Background

Accurate and statistically reliable estimates on the health of California's diverse population of more than 38 million (U.S. Census Bureau, 2014) are critical to developing public health policy for the state. The University of California, Los Angeles (UCLA) has partnered with the Department of Health Care Services, the California Department of Public Health, and RTI International (RTI) to conduct the California Health Interview Survey (CHIS) 2015-2016 to provide these important data.

Obtaining such vital estimates requires a methodologically rigorous survey design that maximizes participation across all geographical areas and population groups in the state. Specifically, the goals of the CHIS 2015-2016 sample design are to produce reliable estimates (1) for the total residential population of California statewide, (2) for the major race/ethnicity groups across California, (3) for important racial subgroups of Vietnamese and Korean, and (4) for the residential population at the county (or grouped county) level. We will accomplish these goals in CHIS 2015-2016 with interviews conducted from a stratified, dual-frame random digit dial (RDD) sample of landline and cell phone numbers.

In this report, RTI presents the draft sampling plan to meet all objectives for CHIS 2015-2016. Section 2 contains the estimation goals for the current study, highlighting specific changes from previous rounds. Section 3 contains a description of the household-level dual-frame design proposed for the selection of landline and cell phone samples. Within-household sampling of adults, children and adolescents is described in Section 4. Section 5 covers information on the respective sampling frames for the dual-frame RDD design. Once selected, we will randomly divide the independently-drawn quarterly samples into release replicates; Section 6 summarizes our approach for releasing the replicates for data collection conducted in collaboration with the Responsive Design and Data Collection teams. Section 7 contains a series of evaluations we will conduct within the first quarter of data collection. We conclude this report with a brief discussion of the resulting base weights, referencing a forthcoming weighting plan for additional information.

2. Estimation Goals for CHIS 2015-2016

We define the criterion required for estimation from CHIS 2015-2016 in the section below. We begin in Section 2.1 with a definition of the target population for households and persons within household. Section 2.2 contains the current analytic objective for the study.

2.1 Target Population

The sample design for CHIS 2015-2016 is summarily described as a stratified two-stage design with a nonresponse follow-up subsample of nonrespondents. The strata are consistent with the 2013-2014 design and are shown in **Table 2.1**.

Table 2-1. Design Strata: CHIS 2015-2016

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

^a Service Planning Areas are substrata of Los Angeles county and the sample will be proportionately allocated among the substrata.

Estimates from CHIS represent the non-institutionalized, residential population in California including adults (ages 18 years and older), children (ages 11 and younger), and adolescents (ages 12-17 years) living in residential households (i.e., non-group quarters) within California. Residential households are randomly chosen either through the landline telephone frame, the cell phone frame, or possibly both. Households without telephone service are excluded from CHIS.² Eligible residences include households, apartments, or mobile homes containing individuals, (multiple and/or) extended families, or multiple unrelated persons provided that they number less than nine. Institutionalized residences (e.g., prisons, jails, juvenile detention facilities, psychiatric hospitals, extended-stay treatment programs, and long-time care), group quarters (those with nine or more unrelated persons), and the homeless population are not eligible for the study.

We will randomly select one adult from among those 18 years of age and older who reside in the eligible residence. This eligible roster includes those temporarily away from the household (e.g., college students living on campus, temporarily hospitalized, away on business). Proxy interviews will be obtained for any randomly chosen adult who is unable to answer the survey questions themselves because of chronic illness or disability.

We will randomly select one child (age 11 years or younger) and one adolescent (12-17 years of age) from among those for whom the selected adult is the parent or legal guardian. Proxy interviews will be collected from a sufficiently knowledgeable adult for the selected child; adolescents will be recruited directly for the study after receiving permission from a parent or legal guardian.

2.2 Analytic Objectives

We summarize the current analytic objectives for CHIS 2015-2016 in **Table 2.2**. The analytic objectives are subject to change with the addition of oversampling within certain geographic areas or domains of interest (see Section 4.1.4).

Overall, CHIS 2015-2016 is expected to yield at least 40,890 completed adult interviews, 40,000 interviews from the base design across 44 design strata plus a 2015-only oversample for Marin County. Per the statement of work corresponding to projections from CHIS 2013-2014, we expect to obtain interviews from approximately 2,583 adolescents within the selected households and proxy interviews for 6,833 children for the main CHIS 2015-2016 sample.

Additional interviews will be generated for any authorized supplemental samples (i.e., oversamples) for certain counties (e.g., Marin) or for certain domains (e.g., Asian adults).

² Estimates from the 2013 National Health Interview Survey suggest that less than 2.0 percent of California households do not have either a landline or cell phone, and are therefore excluded from sampling for CHIS 2015-2016 (http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless_state_201412.pdf).

For clarity of the CHIS design, this sampling plan only addresses supplemental samples that are currently part of the primary CHIS contract (e.g., and Korean and Vietnamese oversamples). The sample design for the Building Healthy Communities (BHC) contract and the association of CHIS and BHC will be discussed in a separate deliverable associated with that contract. The desired split by landline and cell phone interviews is 50-50 as noted in Table 2.2. This ratio is different from CHIS 2013-2014 where approximately 80% of the interviews were obtained from the landline telephone frame.

Table 2-2. Goals for the Minimum Number of Completed Interviews by Sample Characteristics to Meet Analytic Objectives: CHIS 2015-2016

Interview Characteristics^a	Target Minimum Number of Completed Interviews (including primary and oversample)
<u>State-wide, Overall</u>	
Adults (18 years of age or older)	40,000
Landline interviews	20,000
Cell Phone interviews	20,000
Adolescents (12-17 years of age)	2,583
Children (0-11 years of age)	6,833
<u>State-wide, Sampling by Racial Subgroups^b</u>	
Adults, Korean	500
Adults, Vietnamese	500

^a Approximately 50% of the interviews ($\pm 5\%$) will be conducted via landline phone.

^b Target interviews for the base CHIS sample plus any supplement samples needed to reach these goals.

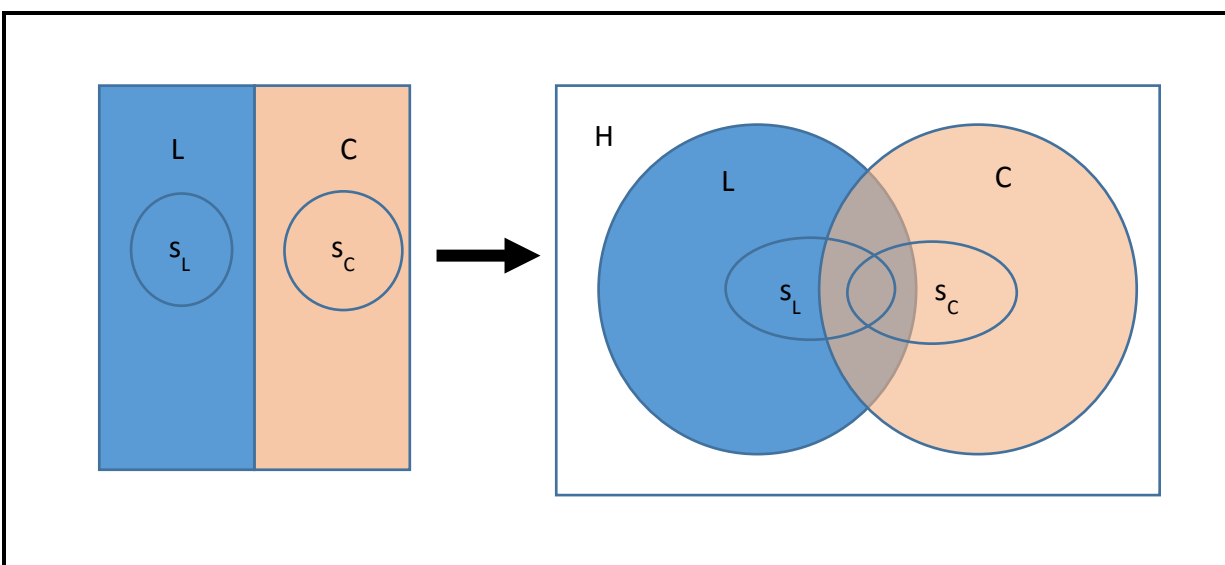
3. Sampling Frames

When CHIS was first conducted, only a list-assisted RDD landline telephone frame was used to generate a telephone sample. However, due to telephone-use changes brought about by the introduction of cell phones, samples from a cell-phone frame have been included with the landline sample since 2007, making the CHIS a dual-frame survey. Today, cell phones are used by more people in California than landlines (Blumberg and Luke, 2014).

Therefore, this dual-frame methodology will be continued in CHIS 2015-2016, with a higher sample allocation to cell phone numbers. In this section, we discuss each sampling frame, along with critical information needed to identify working phone numbers and subgroups requiring oversampling.

The CHIS uses a dual-frame strategy that allows two telephone frames to represent the household population within each stratum, as shown in **Figure 3-1**. This figure shows the goal of the sample design: to map samples drawn from the telephone frames to a sample of eligible households. The left-hand side shows the two mutually-exclusive telephone frames, the landline frame L (in blue) and the cell phone frame C (in pink), for California. These frames contain all of the residential phone numbers that are eligible for selection. Landline numbers may be ported to cell phones (and potentially vice versa) but the number is not part of both frames. Therefore, these frames facilitate the selection of independent samples.

The right-hand side of Figure 3-1 represents the households identified through the landline and cell phone frames. For each stratum, a sample of telephone numbers, s_L , is drawn from L and a sample of numbers, s_C , is drawn from C. Note however, that some households may be contacted through at least one unique landline phone and at least one unique cell phone as represented by the intersection of the sample sets. These people are members of the “dual use” population because they are on both frames.

Figure 3-1. Transformation of Telephone Frames to Household Population

Sizes are not representative of the true frame and sample sizes and frame overlap.

3.1 Landline Frame

The landline frame consists of all working 100-number banks. A 100-number bank contains 100 landline numbers with the same first 8 digits (e.g., 310-794-2800 through 310-794-2899), and each 100-number bank is included in the frame if it has at least one working residential number that is matched to a listing in a public directory. The banks are treated as clusters, and this clustered design improves operational efficiency by focusing the selection of numbers from 100-number banks with at least one working residential number. Furthermore, this design allows for the selection of unlisted numbers because of clustering of residential numbers within a 100-number bank.

The landline sample will be purchased from Marketing Systems Group (M-S-G), the vendor that has provided landline RDD sample for the previous rounds of CHIS. All numbers will be delivered to RTI regardless of their pre-screened status for further evaluation (see Section 4). Also in Section 4, we include an evaluation of sample from two vendors—M-S-G and Survey Sampling International (SSI) in the first quarter of data collection.

The population counts for the landline frame obtained from M-S-G in January 2015 are shown in **Table 3-1**. Population counts from SSI are forthcoming. The Estimated Residential Rate is the percent of numbers in working 100-blocks (i.e., those with at least one residential number) that are estimated to be residential. For example, for Los Angeles County, there are an estimated 2,370,000 working residential numbers.

Table 3-1. M-S-G Landline Frame Counts and Estimated Residential Rates (January 2015) by Stratum: CHIS 2015-2016

Stratum	Working 100-Blocks	Estimated Residential Rate	Stratum	Working 100-Blocks	Estimated Residential Rate
1 – Los Angeles	87,776	27%	23 – San Luis Obispo	2,562	30%
2 – San Diego	27,434	28%	24 – Placer	3,441	34%
3 – Orange	29,389	25%	25 – Merced	1,373	33%
4 – Santa Clara	16,952	26%	26 – Butte	1,799	37%
5 – San Bernardino	14,096	32%	27 – Shasta	1,586	34%
6 – Riverside	14,826	35%	28 – Yolo	1,508	29%
7 – Alameda	15,896	27%	29 – El Dorado	1,706	34%
8 – Sacramento	12,468	29%	30 – Imperial	1,006	31%
9 – Contra Costa	10,282	30%	31 – Napa	1,233	32%
10 – Fresno	6,807	30%	32 – Kings	755	35%
11 – San Francisco	11,274	23%	33 – Madera	897	34%
12 – Ventura	6,848	30%	34 – Monterey	3,685	26%
13 – San Mateo	8,559	26%	35 – Humboldt	1,334	29%
14 – Kern	5,489	28%	36 – Nevada	1,107	37%
15 – San Joaquin	4,540	34%	37 – Mendocino	873	31%
16 – Sonoma	4,816	34%	38 – Sutter	663	38%
17 – Stanislaus	3,652	34%	39 – Yuba	557	29%
18 – Santa Barbara	3,897	27%	40 – Lake	677	30%
19 – Solano	3,345	36%	41 – San Benito	378	31%
20 – Tulare	2,730	31%	42 – Tehama-...	868	32%
21 – Santa Cruz	2,806	29%	43 – Del Norte-...	1,832	23%
22 – Marin	3,555	30%	44 – Tuolumne-...	2,438	31%

3.2 Cell Phone Frame

The cell phone frame was introduced to the CHIS in 2009 to accommodate changes in telephone use among the population. Blumberg and Luke (2014) estimated that 36.4% of the California households had become cell-phone-only households in 2014. This presents a problem with only using a landline-based sample in that socioeconomic and demographic differences exist between adults in cell-phone-only houses versus landline-only houses. Consequently, sampling from only a landline frame will introduce a large bias in the final estimates.

January 2015 counts from the M-S-G cell phone frame for rate centers in California are shown in **Table 3-2**. Rate centers were mapped to CHIS 2015-2016 design strata by county FIPS code. The cell phone frame is comprised of dedicated 1000-number blocks. Unlike the landline frame (Table 3.1), however, rates of working residential numbers for the cell phone frame are unknown. Numbers ported from landline to cell phone status will be identified just before data collection as discussed in Section 4. Population counts from SSI's cell phone frame are forthcoming.

Table 3-2. M-S-G Cell Phone Frame Counts by Stratum: CHIS 2015-2016

Stratum	1000- Number Blocks	Stratum	1000- Number Blocks
1 – Los Angeles	27,710	23 – San Luis Obispo	331
2 – San Diego	4,393	24 – Placer	1,768
3 – Orange	8,584	25 – Merced	257
4 – Santa Clara	5,192	26 – Butte	266
5 – San Bernardino	5,074	27 – Shasta	233
6 – Riverside	5,110	28 – Yolo	676
7 – Alameda	11,985	29 – El Dorado	572
8 – Sacramento	7,296	30 – Imperial	298
9 – Contra Costa	5,435	31 – Napa	127
10 – Fresno	1,252	32 – Kings	146
11 – San Francisco	8,040	33 – Madera	151
12 – Ventura	1,033	34 – Monterey	521
13 – San Mateo	3,930	35 – Humboldt	167
14 – Kern	1,052	36 – Nevada	97
15 – San Joaquin	777	37 – Mendocino	116
16 – Sonoma	606	38 – Sutter	1,616
17 – Stanislaus	624	39 – Yuba	0
18 – Santa Barbara	535	40 – Lake	56
19 – Solano	493	41 – San Benito	134
20 – Tulare	456	42 – Tehama-...	56
21 – Santa Cruz	296	43 – Del Norte-...	67
22 – Marin	1,595	44 – Tuolumne-...	52

3.3 Supplemental Surname Lists

CHIS has used two approaches to oversample Vietnamese and Koreans: 1) oversampling telephone exchanges within counties that have higher Korean and Vietnamese population density (high-density strata), and 2) using surname frames. We plan to use the first approach, (i.e., higher sampling rate in high-density strata). On the Ohio Medicaid Assessment Survey (OMAS), we found that where we targeted minority populations using a surname frame in addition to landline and cell phone frames the cell phone samples substantially increased the proportion of minorities captured outside of the surname frame (unpublished). Owing to its limited usefulness, surname lists are no longer used in the OMAS sample design.

There are several weaknesses to the surname frame approach. First, they only include telephone numbers from the listed landline frame that are flagged as having a surname that is common in a particular racial or ethnic group, and exclude even common names if they are also common in another ethnic group. Second, surname frames cannot capture the racial/ethnic population that is unlisted. Third, when combined with RDD data, surname samples can substantially increase design effects because they contain a relatively small number of landline numbers and must be substantially oversampled to achieve the desired targets.

Strata containing a relatively high proportion of Koreans or Vietnamese were identified using the 2013 American Community Survey 5-Year estimates³. We will adjust the use of high-density strata and even the allocation to cell phones to reach the ethnic goals without the use of surname lists based on results from the quarter of data collection if we project the need and only after consulting with UCLA. We will retain the ability to use the Vietnamese and Korean surname frames starting later in 2015, if needed.

³ <http://www.census.gov/acs/www/>

4. Household Sampling Design

The CHIS 2015-2016 sample design retains much of its character from the 2013-2014 study. Namely, CHIS 2015-2016 is a stratified, two-stage dual-frame design where study-eligible households are contacted either through a landline or cell phone number in the first stage, and one adult resident of the household (18 years of age or older) is chosen at the second stage. Additionally, if the randomly chosen adult is the biological parent or legal guardian of a child (0-11 years of age) or an adolescent (12-17 years of age), then additional subsampling will occur for those less than 18 years of age. There are, however, a few changes to the CHIS design.

New to the 2015-2016 design is the increase in the size of the cell phone sample from an 80-20 landline-cell phone split in 2013-2014 to yield a 50-50 split in the number of completed interviews. Benefits of such a change include access to a growing percent of the California residential population that have a cell phone or are living in a cell-phone only household, and with that likely access to important racial subgroups (Blumberg and Luke, 2014).

Also new is the inclusion of a nonresponse follow-up study where a second phase of sampling is used to target certain nonrespondents. We refer to the initial sampling as the phase-1 design and the nonresponse subsampling as the phase-2 design. Details of the sampling design and selection for phase 1 are provided below (Section 4.1). Features of the phase-2 design and selection are summarized in Section 4.2 below and detailed in the forthcoming Responsive Design plan.

Table 4.1 contains a summary of the projected sample sizes required to meet the analytic objectives for CHIS 2015-2016 included in our response to the revised statement of work. The assumptions are based on the culmination of prior work such as the National Intimate Partner and Sexual Violence Surveillance System, and will be revised as needed based on current information from CHIS 2015-2016. The estimates exclude any proposed supplemental sampling of cases discussed in Section 4.1.4. The sample will be purchased in seven quarterly samples—three in 2015 and four in 2016.

Table 4-1. Projected Sample Purchased and Released, and Expected Number of Completed Interviews by Frame, Interview Type, and Design Phase: CHIS 2015-2016

Frame by Interview	Sample			Phase 1		Phase 2		Overall	
	Purchased	Screened	Released	Interviews	Response Rate ^a	Interviews	Response Rate ^a	Interviews	Response Rate ^{a,b}
<u>Landline</u>	471,649	210,231	191,119						
Screener				36,998	17.0%	2,252	10.9%	39,250	26.0%
Adult				17,600	15.0%	2,400	9.6%	20,000	23.2%
Child				3,317	15.2%	217	9.7%	3,534	23.4%
Adolescent				1,282	7.4%	87	4.7%	1,369	11.7%
<u>Cell Phone</u>	288,901	288,901	262,638						
Screener				34,565	18.3%	1,718	6.9%	36,283	24.0%
Adult				17,600	17.8%	2,400	6.7%	20,000	23.3%
Child				3,809	14.3%	176	5.4%	3,985	18.9%
Adolescent				1,386	8.4%	87	3.2%	1,473	11.3%

^aThe overall combined response rate is calculated as $(1 - [1 - \text{Phase 1 RR}] * [1 - \text{Phase 2 RR}])$. Because of the increased effort in Phase 2, the Phase 2 response rate is higher than what would be obtained if the same protocol was followed during those contacts. Thus, while the Phase 2 response rate is lower than the Phase 1 response rate, combining the two rates produces an overall higher combined value. The phase 2 sampling rate of 40% assumed may be adjusted based on information from the quarterly data collection and budget constraints.

4.1 Phase-1 Sample Design and Selection

We will purchase landline and cell phone samples from an approved vendor four times a year based on detailed specifications provided by RTI. For the first quarter of data collection in 2015, RTI will purchase sample from two vendors (M-S-G and SSI) to assess beneficial differences in their sampling frames and screening procedures for CHIS. We intend to choose a single vendor for the remaining quarters of CHIS based on a comparative analysis (see Section 7.2 for more details). Selecting sample just before data collection ensures that any appended data contain the most up-to-date information. We discuss the nuances of each sample below.

4.1.1 Landline Sample

The vendor will select a stratified simple random sample of landline telephone numbers from the frame of working 100-number blocks discussed in the previous section. Assignment of telephone numbers to the geographic design strata displayed in Table 2.1 is based on area code. **Table 4.2** displays the latest landline frame counts from M-S-G by stratum, along with the estimated number of sample cases requested for CHIS 2015-2016 and the minimum desired number of completed adult interviews. The approximate inflation rate from completed questionnaires to the sample is 23.6 based on previous experience as discussed above, and accounts for sample loss during data collection (e.g., ineligible numbers, ineligible households, and nonresponse) plus additional sample that may not be released.

Table 4-2. M-S-G Landline Frame Counts (January 2015), Expected Total Sample Size, and Desired Number of Completed Adult Interviews by Stratum: CHIS 2015-2016

Stratum	Landline Frame	Expected Total Sample Size	Completed Adult Interviews ^a
1 – Los Angeles	8,777,600	91,028	3,860
2 – San Diego	2,743,400	37,024	1,570
3 – Orange	2,938,900	26,271	1,114
4 – Santa Clara	1,695,200	17,923	760
5 – San Bernardino	1,409,600	15,588	661
6 – Riverside	1,482,600	24,290	1,030
7 – Alameda	1,589,600	14,267	605
8 – Sacramento	1,246,800	16,366	694
9 – Contra Costa	1,028,200	12,428	527
10 – Fresno	680,700	10,589	449
11 – San Francisco	1,127,400	8,843	375
12 – Ventura	684,800	6,745	286
13 – San Mateo	855,900	8,277	351
14 – Kern	548,900	8,678	368
15 – San Joaquin	454,000	5,896	250
16 – Sonoma	481,600	5,896	250
17 – Stanislaus	365,200	5,896	250
18 – Santa Barbara	389,700	5,896	250
19 – Solano	334,500	5,896	250
20 – Tulare	273,000	5,896	250
21 – Santa Cruz	280,600	5,896	250
22 – Marin (2015 oversample)	355,500	16,390	695
23 – San Luis Obispo	256,200	5,896	250
24 – Placer	344,100	5,896	250
25 – Merced	137,300	5,896	250
26 – Butte	179,900	5,896	250
27 – Shasta	158,600	5,896	250
28 – Yolo	150,800	5,896	250
29 – El Dorado	170,600	5,896	250
30 – Imperial	100,600	5,896	250
31 – Napa	123,300	5,896	250
32 – Kings	75,500	5,896	250
33 – Madera	89,700	5,896	250
34 – Monterey	368,500	5,896	250
35 – Humboldt	133,400	5,896	250
36 – Nevada	110,700	5,896	250
37 – Mendocino	87,300	5,896	250
38 – Sutter	66,300	5,896	250
39 – Yuba	55,700	5,896	250
40 – Lake	67,700	5,896	250
41 – San Benito	37,800	5,896	250
42 – Tehama-Glenn-Colusa	86,800	4,716	200
43 – Del Norte-Siskiyou-Lassen-Trinity-Modoc-Plumas-Sierra	183,200	4,716	200
44 – Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	243,800	4,716	200
TOTAL	32,971,500	482,143	20,445

^a *The overall unequal weighting effect (i.e., design effect of the expected base weights) is 1.19 for the landline sample, which does not include coverage and nonresponse adjustments.*

Per discussions with M-S-G, supplemental information is available only for listed landline phones number. Vendors obtain this information from address-based sampling frames, by way of statistical models, and through additional sources. The information includes:

- for each member of a household with at most six adults, name, gender, age, model-estimated race/ethnicity, model-estimated religion, model-estimated occupation (modeled), Hispanic surname flag, Asian surname flag, and marital status;
- number of adult residents;
- number of children;
- model-estimated indicator of children in the household by age group (0-3 years, 4-6 years, 7-9 years, 10-12 years, 13-15 years, and 16-18 years);
- model-estimated household income;
- home-owner status;
- type of dwelling unit; and
- single versus multiple families in residence.

Note that the accuracy of the supplemental information is unknown at this time. Because of this and the availability only for listed numbers, we will not include this information as part of the sampling design at this point. We will request the supplemental information for all sample cases and evaluate its utility each quarter for CHIS 2015-2016 data collection in future quarters.

The vendor will pre-screen phone numbers prior to delivery and include on the files indicators for business and non-working status. Over the first quarters of data collection, we will conduct a series of tests to compare M-S-G provided non-working and business information against information obtained from the RTI dialer. The RTI dialer has proved beneficial to screen out nonworking and business numbers prior to the sample being fielded by interviewers for OMAS data collection (unpublished). We will not field known business phone numbers for CHIS 2015-2016. We will document the benefits and any drawbacks from this approach and discuss with UCLA before modifying the sampling plan to field known business numbers.

We will additionally screen the sample cases in house for instances where landline numbers were ported to cell phone. This step will be conducted just prior to fielding to ensure the most current status. Landlines ported to cell phones will be routed through the CHIS 2015-2016 screener as if they were originally selected from the cell phone frame (and vice versa).

During the first quarter of data collection, RTI will conduct a second set of comparisons by sample vendor. We will request half of the sample from M-S-G and the remaining half from Survey Sampling International (SSI). Though proprietary, our past experience indicates that not only are the telephone sampling frames defined differently among these two vendors (e.g., varying list of working number-blocks), but the procedures for identifying non-working numbers differs as well (Peytchev & Krotki, 2010). Given the time that has elapsed since this initial evaluation, we suggest that it is worth examining whether these differences are still significant. As with the dialer evaluation, we will document the findings from our comparison study and discuss with UCLA before modifying the sampling plan for future sample releases.

4.1.2 Cell Phone Sample

The same vendor will also select a stratified simple random sample of cell phone numbers from the frame of working 1000-number blocks discussed in the previous section. Assignment of telephone numbers to the geographic design strata displayed in Table 2.1 is rate center. **Table 4.3** displays the latest cell phone frame counts from M-S-G by stratum, along with the estimated number of sample cases requested for CHIS 2015-2016 and the minimum desired number of completed adult interviews. The approximate inflation rate from completed questionnaires to the sample is 14.4 based on previous experience as discussed above, and accounts for sample loss during screening and data collection plus additional sample that may not be released.

Table 4-3. M-S-G Cell Phone Frame Counts (January 2015), Expected Total Sample Size, and Desired Number of Completed Adult Interviews by Stratum: CHIS 2015-2016

Stratum ^a	Cell Phone Frame	Expected Total Sample Size	Completed Adult Interviews ^b
1 – Los Angeles	27,710,000	55,758	3,860
2 – San Diego*	4,393,000	22,679	1,570
3 – Orange	8,584,000	16,092	1,114
4 – Santa Clara	5,192,000	10,978	760
5 – San Bernardino*	5,074,000	9,548	661
6 – Riverside	5,110,000	14,878	1,030
7 – Alameda*	11,985,000	8,739	605
8 – Sacramento	7,296,000	10,025	694
9 – Contra Costa*	5,435,000	7,613	527
10 – Fresno	1,252,000	6,486	449
11 - San Francisco	8,040,000	5,417	375
12 – Ventura	1,033,000	4,131	286
13 - San Mateo	3,930,000	5,070	351
14 – Kern	1,052,000	5,316	368
15 – San Joaquin	777,000	3,611	250
16 – Sonoma	606,000	3,611	250
17 – Stanislaus	624,000	3,611	250
18 – Santa Barbara	535,000	3,611	250
19 – Solano	493,000	3,611	250
20 – Tulare	456,000	3,611	250
21 – Santa Cruz	296,000	3,611	250
22 – Marin (2015 oversample)	1,595,000	10,039	695
23 – San Luis Obispo	331,000	3,611	250
24 – Placer	1,768,000	3,611	250
25 – Merced	257,000	3,611	250
26 – Butte	266,000	3,611	250
27 – Shasta	233,000	3,611	250
28 – Yolo	676,000	3,611	250
29 – El Dorado	572,000	3,611	250
30 – Imperial*	298,000	3,611	250
31 – Napa	127,000	3,611	250
32 – Kings	146,000	3,611	250
33 – Madera	151,000	3,611	250
34 – Monterey	521,000	3,611	250
35 – Humboldt	167,000	3,611	250
36 – Nevada	97,000	3,611	250
37 – Mendocino	116,000	3,611	250
38 – Sutter	1,616,000	3,611	250
39 – Yuba ^c	0	3,611	250
40 – Lake	56,000	3,611	250
41 – San Benito	134,000	3,611	250
42 – Tehama-Glenn-Colusa	56,000	2,889	200
43 – Del Norte- ...	67,000	2,889	200
44 – Tuolumne- ...	52,000	2,889	200
TOTAL	109,175,000	295,330	20,445

^a Starred strata are those where at least 10% of the cell phones were associated with responding adults living in another location based on unpublished CHIS 2013 statistics.

^b The overall unequal weighting effect (i.e., design effect of the expected base weights) is 1.74, which does not include coverage and nonresponse adjustments.

^c Yuba county does not currently have a rate center.

As noted for the landline sample, supplemental information is available from M-S-G but only for a small portion of cell phone numbers, i.e., those with identifying information (i.e., “listed”), and will not be used for the sample design. We will request the supplemental information for all sample cases and evaluate its utility for CHIS 2015-2016 data collection.

The vendor will pre-screen the sample prior to release for data collection and include on the files the active status (active, inactive, or unknown) indicators for the cell numbers (i.e., M-S-G Cell-Wins flag). These flags will be appended as close to replicate release as possible to ensure that the information obtained is current. Over the first quarters of data collection, we will conduct a series of tests to compare M-S-G provided non-working information against information obtained from the RTI dialer. We will document the benefits and any drawbacks from this approach and discuss with UCLA before modifying the sampling plan for future sample releases.

During the first quarter of data collection, RTI will conduct a second set of comparisons by sample vendor as discussed for the landline sample. We will request half of the cell phone sample from M-S-G and the remaining half from SSI. As with the dialer evaluation, we will document the findings from our comparison study and discuss with UCLA before modifying the sampling plan or choice of vendor for future sample releases.

4.1.3 Oversampling for Racial Subgroups

A sampling goal for CHIS 2015-2016 of particular note is the completion of 500 interviews each with (self-reported) Vietnamese and Korean adult residents of California. Two-year projections conducted by UCLA based on unpublished results from CHIS 2013-2014 suggest that under the current 50-50 CHIS 2015-2016 phase-1 design approximately 460 Vietnamese and 359 Korean completed adult interviews will be obtained. However, new to CHIS in 2015-2016 is the inclusion of a nonresponse follow-up phase (i.e., phase-2 subsampling) and phase-2 incentives which we anticipate will benefit the completion rate for the racial groups. Additionally, we will investigate the need for oversampling in the following strata towards the statewide goal of 500 completes: (3) Orange, (4) Santa Clara, (11) San Francisco, and possibly (1) Los Angeles. Results from the first and possibly second quarters of data collection for the two-phase 50-50 design will further inform the level of oversampling.

4.1.4 Supplemental Samples

This section discusses county or sub-group supplemental samples that will be added to the primary CHIS sample sporadically throughout the data collection cycle as requests are made and contracts settled. These supplemental samples are divided into “Committed” and “Exploring” depending on the level of commitment from the sample sponsor and stage in the contracting process. To date, no contracts have been finalized for additional sample. Other supplemental samples are under consideration. Most require marginal costs estimated from the first quarter of data collection to finalize.

COMMITTED – Contract (almost) established between UCLA and Sample Funder

Supplemental Sample #1: Marin County

As of the deliverable date of this plan (March 26, 2015), a contract for supplemental sampling is currently underway for Marin County (stratum 22). The additional interviews are desired only in 2015 as shown in **Table 4-4**.

Table 4-4. Base and Supplemental Sample Completed Interviews Targets for Marin County, California: CHIS 2015-2016

Interview	Base sample Targets		2015 Supplemental Sample Target			CHIS 2015-2016		
	2015	2016	Child in HH	Hispanic	Total	2015	2016	Total
Adult	250	250	780	110	890	1,140	250	1,390
Child	35	35	275	20	295	330	35	365
Teen	15	15	110	10	120	135	15	150

Supplemental Sample #2: Japanese Oversample (Keiro Foundation)

The Keiro Foundation is likely to fund a statewide supplemental sample of California adult residents of Japanese descent to achieve 500 completed interviews across the two years of CHIS 2015-2016. Previous studies used a surname list with only a landline sample. Use of this approach for the CHIS 2015-2016 Japanese oversample must be evaluated especially in light of recent findings that suggest surname lists have limited scope (see, e.g., unpublished findings from OMAS).

Supplemental Sample #3: Building Healthy Communities (BHC, The California Endowment)

The California Endowment will fund a supplemental sample of interviews in at most 14 communities within California. RTI will provide details of the sampling plan, intended to focus on an addressed-based sampling (ABS) methodology using RTI’s enhanced ABS frame, in a forthcoming report.

EXPLORING – Request Made or Anticipated from Sponsor

Most Likely

- San Diego County (stratum 2) – supplemental samples to produce an equal allocation across the Health Service Regions (HSR) similar to last cycle (700 each across 6 HSRs).
- Statewide supplemental sample of African Americans and Asians used by Covered California to “stabilize subgroup estimates for a simulation of California health insurance markets (dubbed “CalSIM” a joint project of our Center and the Berkeley Labor Center)” (communique from David Grant, UCLA on Jan 28, 2015).

Least Likely

- Ventura County (stratum 12)
- Los Angeles County (stratum 1) – supplemental sample of possibly 1,200 households with adolescents for a teen-focused interview.

4.2 Phase 2 Sample Selection

A nonresponse follow-up (phase 2) subsample will be selected for CHIS 2015-2016 based on analyses conducted during data collection using paradata and any available screener responses. A stratified sample of nonrespondents in each design stratum will be selected to minimize errors while controlling costs. The original proposal recommended at 40% sub-sampling rate at the second phase, with an increased incentives of \$5 for a completed screener, \$25 for a completed adult interview, \$10 for completed child interviews, and \$10 for completed teen interviews. Additional information on the responsive design plan is provided in a forthcoming memo, and nonresponse recruitment materials can be found as part of Deliverable 4.4.

5. Within-Household Sampling

In this section, we describe the methods for selecting CHIS study participants from within a sampled household. We will capture data from one randomly chosen adult resident. If present and eligible for the study, we will also randomly select one adolescent to be directly interviewed, and one child whose data will be collected by a sufficiently knowledgeable adult who is also the parent of that child. The probability of selection of an individual within a household differs between households because of the household composition. Therefore, we will use different and, in some cases, multiple selection procedures to select individuals for the survey.

We first discuss the procedure for selecting an adult from each household (Section 5.1). Next, we outline the selection procedure for children (Section 5.2.1). Finally, the procedure for selecting an adolescent for the interview is presented (Section 5.2.2). Throughout the data collection process, we will monitor screening rates and the number of completed interviews from each of the target populations (adults, adolescents and children) to ensure that the selection methods are obtaining the needed individuals to meet sample size targets.

5.1 Adult Sampling and Eligibility

Adults are any person who is 18 years of age or older. Because of the dual frames, the method of selecting an adult will differ slightly based on whether the number is from the landline frame or the cell phone frame. For the landline sample, the Rizzo method of selection will be used to select an adult from a sampled household (Rizzo et al., 2004). This is a modified next-birthday method that does not require enumerating all adults within a household, thereby reducing screener duration and respondent burden while giving each adult an equal probability of selection within a selected household. The outline of the method follows:

1. During the screening interview, the interviewer will ask the contacted individual for the number of adults in the household.
2. If only one adult lives in the household, then that adult is selected.
3. If two adults live in the household, each adult has a 50% chance of being selected. In practice, the computer assisted telephone interview (CATI) system will generate a random number between 0 and 1. If the number is less than or equal to 0.5, then the contacted adult is selected for the interview. Otherwise, the other adult is selected for the interview.
4. If more than two adults live in the household, then a more detailed procedure must be implemented that assigns a probability of selection to each person as the inverse

of the number of people in the household, and that selects an individual without requiring the screener respondent to list all the people in the household. This is done with the following steps:

- a. The CATI system will randomly generate a number between 0 and 1.
 - b. If that number is less than the inverse of the number of adults in the household, then the contacted person is selected.
 - c. Otherwise, the interviewer will ask the contacted person for the name of the adult that will have the next birthday in the household.
 - i. If the contacted person (i.e., screener respondent) knows which adult that is, that person is selected for the interview and the interviewer will ask to speak with that person.
 - ii. If the screener respondent can (will) not answer that question, the interviewer will request the names of all other adult residents in the household and the CATI system will randomly select one of these individuals.
5. If the respondent does not know the number of adults in the house, the interviewer will request the names of all other adults in the household. The CATI system will randomly select one of these individuals.

We assume that the adult answering the sampled cell phone number is the primary user.

5.2 Selection of Child and Adolescent from Legal Guardian

Children and adolescents under 18 years of age are selected for CHIS 2015-2016 using a two-phase process. First, an adult is selected in a sampled household using the methods in the previous section. One child and one adolescent are selected using methods outlined in the next two sections.

In an effort to increase the rate of obtaining interviews from children and adolescents, researchers for CHIS 2005 introduced a method of selecting children and adolescents, called the child-first method. This method allowed interviewers to obtain completed interviews for children and consent from adolescents by proxy if the selected adult was the spouse or partner to the screener respondent on the landline frame. The structure of the screener directs the interviewer to inquire about the relationship status of the selected adult to the respondent if the respondent is not the selected adult and children are associated with the selected adult. If the respondent is related to the selected adult, he or she can consent to a child interview and give permission for the adolescent interview without waiting for the selected adult to respond. During the 2011-2012 data collection period, 65.9% of the households in which the child-first method was used completed the child extended interview

whereas only 17.3% of the households in which the child-first method was not used completed the child extended interview. Furthermore, adolescent extended interviews were obtained from 31.8% of the households in which the child-first method was implemented versus only 6.6% of the households in which the child-first methods was not used. Because the adult- and child-extended interviews might not occur at the same time, it is possible to obtain data on the selected child and/or adolescent without obtaining the data from the selected adult, thus the child-first questionnaire contains essential information for weighting that is otherwise included in the adult interview. The child-first method was not implemented on the cell frame because the screener respondent was almost always the selected adult.

5.2.1 Child Sampling

A child is defined for CHIS as a person less than 12 years of age normally residing in the eligible household. Eligible children are those who are the legal child of the sampled adult, which excludes certain children who may live in the household, such as foster children. As in previous iterations of the CHIS, children will be sampled at different rates based on their ages. Children five years of age and younger associated with the selected adult will be sampled at twice the rate as the older children ages 6-11 years. The probability of selecting a child in the 0-5 year groups is defined as $2n_{1ij} / (2n_{1ij} + n_{2ij})$ where n_{1ij} is the number of eligible children ages 0-5 years and n_{2ij} is the number of children ages 6-11 years within household i . The corresponding selection probability for eligible children ages 6-11 years is $n_{2ij} / (2n_{1ij} + n_{2ij})$. The screener respondent or sample adult respondent answer questions about the sampled child.

5.2.2 Adolescent Sampling

An adolescent is defined for CHIS as a person between the ages of 12 and 17 years normally residing in the eligible household. Like the child, they must also be the legal child of the selected sample adult. Each adolescent for whom the selected adult is the parent or legal guardian has an equal probability of selection. The selection of one adolescent to participate in CHIS 2015-2016 occurs in one location during the interview process. Whether or not a child-first method is used, adolescents are enumerated at the end of the screener interview in the screener interview. In both cases, one adolescent is selected at random for the extended adolescent interview from the enumerated adolescents. If only one associated adolescent exists in that household, that adolescent is selected. The sampled adolescent is interviewed directly after parent permission is obtained and the teen agrees to participate.

6. Plans for Sample Release

New to CHIS 2015-2016 is the selection and release of quarterly samples instead of the 6-month samples of the past. This ensures updated information for the sampled numbers and allows sufficient time to fully work the selected telephone numbers, paying particular attention to phase-1 refusal conversion to lower nonresponse bias. As discussed in Section 4 of this report, we will select extra phone numbers with each quarterly sample and only field sample needed to meet the study goals.

The plan is to select relatively equal-sized samples for each of the three quarters in 2015 to achieve the desired number of completed interviews. We also plan to select relatively equal-sized samples for the four quarters of 2016. This approach helps to maintain a consistent workload for the interviewers as well as interviews conducted throughout the calendar year.

Each quarterly sample of landline and cell phone numbers will be processed upon receipt from the vendor. In the least, these procedures include:

- confirmation of the sample counts by frame against the specifications document,
- examination of the appended information including business and non-working status,
- elimination of any duplicate phone numbers within and across the quarterly samples,
- creation of a unique record identifier,
- identification of telephone numbers ported from landline to cell phone, and
- construction of within-quarter data release replicates.

We will randomly divide the sampled landline and cell phone numbers each into data-release replicates within the design strata. Through the random assignment all replicates released for data collection are considered a random subsample from the original full quarterly sample. The weights are adjusted for any subsampling as discussed in the forthcoming weighting plan.

Replicates that are too large hinder our ability to release sample cases to meet specific goals (e.g., number of completed interviews by racial subgroup) without exceeding project resources and potentially decreasing response rates. Replicates that are too small only require additional procedures to monitor their status prior to and during data collection. Hence, the latter condition is preferred. For the first quarter of CHIS 2015-2016 data collection, we will follow the lead of the OMAS team and create relatively equal-sized replicates within frame type (landline, cell) and design stratum: landline replicates will contain approximately 100 numbers, and cell phone replicates will contain approximately 50

numbers. Replicates will be released on a flow basis to ensure current screening by the RTI dialer and consistent workload for the interviewers. As discussed in the forthcoming Responsive Design plan, the released samples will be closely monitored to identify as early as possible in the quarter when additional sample is needed. We will adjust the replicate sizes and the initial sample release as needed for subsequent quarters of data collection to ensure efficient data collection operations. We will confer with UCLA on all proposed changes to the sampling plan prior to implementation.

Intended dates for the sample preparation steps are found in the project schedule Gantt chart.

7. First Quarter Evaluations And ongoing reports

We plan to conduct a series of evaluations in the first and possibly second quarters of data collection for CHIS 2015-2016. We will confer with UCLA on our interim and final recommendations from the evaluations, along with any proposed changes prior to implementation.

7.1 Study Rates

The sample sizes selected to achieve the desired number of completed interviews (Table 2.2) in this report were calculated based on a set of assumptions for sample loss. We will closely monitor the rates with the addition of each quarter of data collection to determine if changes to the assumptions and hence the sampling plan are needed.

7.2 Sample Vendors

As discussed in Section 4, half of the sample for the first quarter of data collection will be obtained from M-S-G and the other half from SSI. If a cost and quality analysis identifies one vendor over the other, then we will discuss the UCLA the option of proceeding with only one source for sampled cases for the remaining quarters of CHIS 2015-2016.

7.3 Oversampling for Racial Subgroups

We included a justification for excluding surnames lists as sampling frames for CHIS 2015-2016 in Section 3 of this report. Therefore, goals by racial subgroups (e.g., Vietnamese, Korean) must be met through the increased use of cell phone cases along with oversampling in strata with a higher representation for these populations. We will closely monitor return rates to quickly maximize opportunities for attaining the racial subgroup goals such as releasing additional replicates for certain strata and modifying the sampling plan for the subsequent quarter of data collection.

7.4 Supplemental Frame Information

As noted in Section 3, we will evaluate supplemental information included with the sample cases to determine their utility for updates to the sampling plan and for data collection. For example, if we determine that the RTI dialer produces high-quality supplemental information, then we may forgo requests for such vendor information (e.g., nonworking or business status) in the future. Status of the sample and results from our evaluation will be included in a memo at the end of each quarter (Deliverable 1.3).

7.5 Quarterly Reports

Upon purchase of each quarterly sample, RTI will submit a memo with the number of cases selected, overall, by sample type (landline vs. cell phone) and by design stratum

(Deliverable 1.2). Additionally, we will include a summary of the supplemental information available for each type of telephone number.

7.6 Weekly Sample Management Reports

Updated information on the status of the sampled cases will be provided to UCLA on a weekly basis (or as needed) through standardized reports via the Web portal (Deliverable 6.1). Among other items we will provide the number and percent distribution of all fielded cases, overall and by strata, along with a series of study rates (e.g., cooperation, eligibility and completion).

8. Base Weights

A base weight is calculated as the inverse of the probability of selection for each unit of sampling within a survey design. These weights provide the foundation for the final survey weights used to analyze the data to produce the desired population estimates and associated levels of precision.

We will construct four base weights for CHIS 2015-2016: one household (telephone) base weight and three within-household base weights (adult, child and adolescent). The household base weight is calculated as

$$\frac{N_{fh}}{n_{fh}} \quad (8.1)$$

where f indexes the sampling frame (i.e., landline or cell), h indexes the design strata, n_{fh} is the size of the sample selected within stratum h from sampling frame f , and N_{fh} is the associated frame count within stratum fh .

The remaining base weights are a function of the household base weight and the inverse within-household selection probabilities. The within-household selection probabilities take the form,

$$\frac{mos_{fhi+}}{mos_{fhij}} \quad (8.2)$$

where i indexes the selected household, mos_{fhij} is the measure of size of the j^{th} person by type (e.g., child) selected within household hi from sampling frame f , and mos_{fhi+} is the sum of all measures of size by person type associated with household fhi . Note for many sampled cell phone numbers, the adult selection probability is one because they do not share their cell phones with other adults in the household. In this case, the (unconditional) adult base weight equals the household base weight. A measure of size notation is used in expression (8.2) to account for oversampling of children ages 0-5 years (see Section 4.2).

We will calculate the household base weights just after receipt of the sample from the contractor. We will immediately address any anomalies in the data identified by our quality control procedures prior to release of the sample to the field. These procedures include verification of the weight structure, and summing the weights to ensure they total the stratum counts.

We will further detail the construction of the CHIS 2015-2016 base weights in the forthcoming CHIS 2015-2016 Weighting Plan. In addition, we will provide documentation of the proposed adjustments applied to the base weights to calculate the final analysis weights, along with additional quality control procedures applied at each stage in the final weight construction.

References

- Blumberg and Luke, 2014. National Health Interview Survey Early Release Program: Modeled estimates (with standard errors) of the percent distribution of household telephone status for adults aged 18 and over, by state: United States, 2013. Found at http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless_state_201412.pdf.
- Peytchev, A., & Krotki, K. (2010). Experiments in cell phone nonresponse. Paper presented at the American Association for Public Opinion Research, Chicago, IL. Presentation available upon request.
- Rizzo, L., Brick, J., & Park, I. (2004). A minimally intrusive method for sampling persons in random digit dialing surveys. *Public Opinion Quarterly*, 68, 267-274.
- U.S. Census Bureau. (2014). State and county quickfacts, California. Retrieved from <http://quickfacts.census.gov/qfd/states/06000.html>

Chapter 3: Additional Sampling Documentation

The description above in Chapter 2 reflects the planned approach to sampling at the start of the CHIS 2015-2016 cycle. Chapter 3 describes some initial results after data collection. A more detailed assessment of the sampling plan and sampling results will be published with the release of 2016 data. Counts of completed interviews below may differ from those in the other methodology reports due to the order in which reports were written, and subsequent changes in total number of completes due to data cleaning and processing.

Sampled phone numbers by sampling frame

Table 3-1 below displays the count and distribution of sampled phone numbers across the various sampling frames in CHIS 2015. About 70 percent of all sampled telephone numbers for CHIS 2015 were landline numbers and nearly 29 percent were cell phone numbers. The remaining 1.4 percent of the 2015 sampled telephone numbers was drawn from Asian surname sample lists for Vietnamese, Korean, and Japanese ethnicities. This table corresponds to Table 3-9 in the *CHIS 2013-2014 Methodology Report #1 – Sample Design*.

Table 3-1. Number of sampled telephone numbers by sample type

Sample Type	Telephone numbers sampled	
	n	pct
Landline Sample	504,680	70.0
Main Landline Sample	470,116	
Geographic Landline Sample ¹	34,564	
Total Cell Sample	206,429	28.6
Main Cell Sample	181,037	
Geographic Cell Sample ¹	25,392	
Total List Sample	10,016	1.4
Korean List Sample	6,604	
Japanese List Sample	484	
Vietnamese List Sample	2,928	
Total	721,125	100

¹ Oversamples for Marin County.

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

Note: Corresponds to CHIS 2013-14 Table 3-9

Sampled phone numbers by sampling frame and stratum

Table 3-2 below displays the number of phone numbers sampled in CHIS 2015 by sampling stratum and frame within stratum. Consistent with Table 1, the sampled mix of telephone numbers for each geographic stratum in CHIS 2015 was similar to the overall mix of landline, cell phone and Asian surname list sample. Notable deviations from this pattern were Yuba County, for which no cell phone numbers were sampled and 12 strata for which no Asian surname list numbers were sampled. This table corresponds to Table A-2 in CHIS 2013-2014 methodology reports

Table 3-2. Number of telephone numbers sampled by sample frame and sampling stratum

Stratum	Landline	Cell Phone	Surname list frame			Total
			Japanese	Korean	Vietnamese	
State	504,680	206,429	484	6,604	2,928	721,125
1 Los Angeles	109,401	34,149	186	2,229	592	146,557
2 San Diego	37,280	14,430	32	291	211	52,244
3 Orange	37,559	9,918	54	770	658	48,959
4 Santa Clara	23,542	7,851	43	644	488	32,568
5 San Bernardino	17,798	5,379	15	193	76	23,461
6 Riverside	23,860	9,100	15	152	73	33,200
7 Alameda	18,904	4,870	26	544	210	24,554
8 Sacramento	16,791	3,773	23	210	131	20,928
9 Contra Costa	9,940	3,508	18	197	63	13,726
10 Fresno	12,365	3,693	13	60	23	16,154
11 San Francisco	10,831	3,011	18	646	174	14,680
12 Ventura	8,057	2,413	10	67	27	10,574
13 San Mateo	7,962	2,992	17	245	53	11,269
14 Kern	5,835	3,193	0	25	11	9,064
15 San Joaquin	5,432	1,922	8	54	46	7,462
16 Sonoma	4,978	1,380	0	37	16	6,411
17 Stanislaus	4,980	2,860	0	25	11	7,876
18 Santa Barbara	4,234	2,080	0	22	8	6,344
19 Solano	5,550	1,682	0	33	14	7,279
20 Tulare	5,300	1,699	0	11	0	7,010
21 Santa Cruz	4,405	2,260	0	0	0	6,665
22 Marin	40,857	30,245	0	0	0	71,102
23 San Luis Obispo	3,211	2,561	0	11	0	5,783
24 Placer	3,642	2,358	0	33	13	6,046

25 Merced	5,370	1,858	0	8	0	7,236
26 Butte	2,706	1,320	0	13	7	4,046
27 Shasta	2,840	2,035	0	9	0	4,884
28 Yolo	3,510	1,657	0	19	6	5,192
29 El Dorado	3,559	2,017	0	0	0	5,576
30 Imperial	5,349	2,658	0	0	0	8,007
31 Napa	5,173	2,707	0	8	0	7,888
32 Kings	6,691	2,594	0	0	0	9,285
33 Madera	4,492	2,294	0	0	0	6,786
34 Monterey	7,256	1,622	6	28	11	8,923
35 Humboldt	2,343	1,362	0	0	0	3,705
36 Nevada	2,949	2,193	0	0	0	5,142
37 Mendocino	2,711	2,211	0	0	0	4,922
38 Sutter	4,076	13,655	0	0	0	17,731
39 Yuba	4,701	0	0	0	0	4,701
40 Lake	2,894	1,747	0	0	0	4,641
41 San Benito	7,032	3,013	0	0	0	10,045
42 Tehama-Glenn-Colusa	2,922	1,430	0	0	0	4,352
43 Del Norte-Siskiyou-Lassen-Trinity-Modoc-Plumas-Sierra	2,675	1,343	0	8	0	4,026
44 Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	2,717	1,386	0	12	6	4,121

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

Note: Corresponds to CHIS 2013-14 Table A-2

Completed adult interviews by sampling frame and self-reported stratum

Table 3-3 below displays the number of completed adult extended interviews by stratum and sampling frame (landline, cell phone, and Asian surname list). The number of adult interviews reflects both the overall designated sample sizes for each stratum and the outcome rates (such as eligibility and cooperation rates) within each stratum. Note that the number of adult interviews is largest for Los Angeles and San Diego, reflecting larger completed interview goals specified for those counties. This table corresponds to Table A-4 in CHIS 2013-2014 methodology reports.

Table 3-3. Number of completed adult interviews by sample type and self-reported stratum

Stratum	Surname list frame					Total
	Landline	Cell Phone	Japanese	Korean	Vietnamese	
State	11,211	9,360	28	311	124	21,034
1 Los Angeles	2,075	1,781	7	97	14	3,974
2 San Diego	836	628	2	17	14	1,497
3 Orange	577	422	4	30	23	1,056
4 Santa Clara	387	363	2	31	23	806
5 San Bernardino	358	313	2	9	4	686
6 Riverside	554	468	1	5	1	1,029
7 Alameda	349	288	3	32	10	682
8 Sacramento	402	353	1	11	7	774
9 Contra Costa	251	226	3	11	4	495
10 Fresno	288	203	0	3	3	497
11 San Francisco	143	253	1	26	9	432
12 Ventura	188	122	1	3	1	315
13 San Mateo	147	152	0	12	3	314
14 Kern	188	165	0	3	1	357
15 San Joaquin	117	116	0	3	1	237
16 Sonoma	154	149	0	6	0	309
17 Stanislaus	117	123	0	3	0	243
18 Santa Barbara	130	103	0	0	0	233
19 Solano	130	100	0	0	0	230
20 Tulare	132	99	0	0	0	231
21 Santa Cruz	114	126	0	0	0	240
22 Marin	743	397	0	0	0	1,140
23 San Luis Obispo	119	110	0	2	0	231
24 Placer	120	117	0	0	0	237
25 Merced	151	100	0	0	1	252
26 Butte	130	145	0	0	0	275

27 Shasta	130	99	0	0	0	229
28 Yolo	123	106	0	2	2	233
29 El Dorado	120	98	0	0	0	218
30 Imperial	168	80	0	0	0	248
31 Napa	141	97	0	0	0	238
32 Kings	139	104	0	0	0	243
33 Madera	144	95	0	0	0	239
34 Monterey	150	125	1	2	2	280
35 Humboldt	115	95	0	0	0	210
36 Nevada	119	101	0	0	0	220
37 Mendocino	118	103	0	0	0	221
38 Sutter	130	227	0	0	0	357
39 Yuba	130	163	0	0	0	293
40 Lake	109	94	0	0	0	203
41 San Benito	156	110	0	0	0	266
42 Tehama-Glenn-Colusa	106	111	0	0	0	217
43 Del Norte-Siskiyou-Lassen-Trinity-Modoc-Plumas-Sierra	107	67	0	2	0	176
44 Tuolumne-Calaveras-Amador-Inyo-Mariposa-Mono-Alpine	106	63	0	1	1	171

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

Note: Corresponds to CHIS 2013-14 Table A-4

Completed child interviews by sampling frame and self-reported stratum

Table 3-4 below displays the distribution of completed child interviews (completed by parents for children aged 0-11 years) by sample frame – landline numbers, cell phone numbers, and Asian surname list numbers – for each stratum. The number of child interviews reflects both the overall designated sample sizes for each stratum, outcome rates (such as eligibility and cooperation rates), and differential patterns of households with children within and across each stratum. This table corresponds to Table A-5 in CHIS 2013-2014 methodology reports.

Table 3-4. Number of completed child interviews by sample type and self-reported stratum

Stratum	Landline	Cell Phone	Surname list frame			Total
			Japanese	Korean	Vietnamese	
State	997	1,124	2	22	12	2,157
1 Los Angeles	184	208	0	7	2	401
2 San Diego	83	72	0	1	1	157
3 Orange	39	49	0	4	3	95
4 Santa Clara	41	35	0	3	3	82
5 San Bernardino	31	39	0	0	0	70
6 Riverside	43	57	0	0	1	101
7 Alameda	24	31	0	0	0	55
8 Sacramento	31	50	0	1	1	83
9 Contra Costa	24	23	0	0	0	47
10 Fresno	33	35	0	0	0	68
11 San Francisco	10	22	0	3	0	35
12 Ventura	22	21	1	0	0	44
13 San Mateo	13	18	0	3	1	35
14 Kern	31	27	0	0	0	58
15 San Joaquin	11	18	0	0	0	29
16 Sonoma	17	19	0	0	0	36
17 Stanislaus	9	19	0	0	0	28
18 Santa Barbara	7	12	0	0	0	19
19 Solano	9	11	0	0	0	20
20 Tulare	15	15	0	0	0	30
21 Santa Cruz	10	12	0	0	0	22
22 Marin	49	30	0	0	0	79
23 San Luis Obispo	4	14	0	0	0	18
24 Placer	3	12	0	0	0	15
25 Merced	16	17	0	0	0	33
26 Butte	11	16	0	0	0	27
27 Shasta	10	13	0	0	0	23
28 Yolo	12	22	0	0	0	34
29 El Dorado	8	12	0	0	0	20

30	Imperial	24	11	0	0	0	35
31	Napa	11	12	0	0	0	23
32	Kings	31	17	0	0	0	48
33	Madera	6	12	0	0	0	18
34	Monterey	18	20	1	0	0	39
35	Humboldt	11	5	0	0	0	16
36	Nevada	6	5	0	0	0	11
37	Mendocino	15	6	0	0	0	21
38	Sutter	18	35	0	0	0	53
39	Yuba	13	21	0	0	0	34
40	Lake	8	8	0	0	0	16
41	San Benito	15	17	0	0	0	32
	Tehama-Glenn-						
42	Colusa	8	12	0	0	0	20
43	Del Norte-Siskiyou-	9	8	0	0	0	17
	Lassen-Trinity-						
	Modoc-Plumas-Sierra						
44	Tuolumne-Calaveras-	4	6	0	0	0	10
	Amador-Inyo-						
	Mariposa-Mono-						
	Alpine						

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

Note: Corresponds to CHIS 2013-14 Table A-5

Completed teen interviews by sampling frame and self-reported stratum

Table 3-5 below displays the distribution of completed teen interviews (completed by youth aged 12 to 17 with parental permission) by sample frame – landline numbers, cell phone numbers, and Asian surname list numbers – for each stratum. The number of teen interviews reflects both the overall sample sizes designated for each stratum, outcome rates (such as eligibility, permission from parent/guardian, and cooperation rates), and differential patterns of households with teens within and across each stratum. This table corresponds to Table A-6 in CHIS 2013-2014 methodology reports.

Table 3-5. Number of completed teen interviews by sample type and self-reported stratum

Stratum	Landline	Cell Phone	Surname list frame			Total
			Japanese	Korean	Vietnamese	
State	367	367	3	10	7	754
1 Los Angeles	82	54	2	3	1	142
2 San Diego	26	25	0	1	0	52
3 Orange	17	17	0	1	1	36
4 Santa Clara	16	8	0	1	1	26
5 San Bernardino	10	13	0	0	1	24
6 Riverside	11	28	1	0	0	40
7 Alameda	15	5	0	0	0	20
8 Sacramento	12	11	0	1	1	25
9 Contra Costa	7	6	0	0	0	13
10 Fresno	9	12	0	0	1	22
11 San Francisco	2	5	0	3	0	10
12 Ventura	10	2	0	0	0	12
13 San Mateo	3	4	0	0	1	8
14 Kern	17	14	0	0	0	31
15 San Joaquin	0	7	0	0	0	7
16 Sonoma	2	5	0	0	0	7
17 Stanislaus	4	6	0	0	0	10
18 Santa Barbara	8	5	0	0	0	13
19 Solano	2	3	0	0	0	5
20 Tulare	6	9	0	0	0	15
21 Santa Cruz	1	6	0	0	0	7
22 Marin	26	11	0	0	0	37
23 San Luis Obispo	4	4	0	0	0	8
24 Placer	1	5	0	0	0	6
25 Merced	5	9	0	0	0	14

26	Butte	3	6	0	0	0	9
27	Shasta	2	4	0	0	0	6
28	Yolo	2	6	0	0	0	8
29	El Dorado	2	2	0	0	0	4
30	Imperial	13	7	0	0	0	20
31	Napa	3	6	0	0	0	9
32	Kings	3	7	0	0	0	10
33	Madera	4	2	0	0	0	6
34	Monterey	7	4	0	0	0	11
35	Humboldt	3	2	0	0	0	5
36	Nevada	7	7	0	0	0	14
37	Mendocino	2	5	0	0	0	7
38	Sutter	5	11	0	0	0	16
39	Yuba	1	10	0	0	0	11
40	Lake	4	2	0	0	0	6
41	San Benito	4	5	0	0	0	9
42	Tehama-Glenn-Colusa	2	3	0	0	0	5
43	Del Norte-Siskiyou-Lassen- Trinity-Modoc-Plumas- Sierra	3	4	0	0	0	7
44	Tuolumne-Calaveras- Amador-Inyo-Mariposa- Mono-Alpine	1	0	0	0	0	1

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

Note: corresponds to CHIS 2013-14 Table A-6