December 13, 2016

CHIS 2015 Methodology Report Series

Report 5

Weighting and Imputation

-Short Report-

TABLE OF CONTENTS: CHIS 2015 METHODOLOGY SHORT REPORT #5

Section **Page**

Chapter 1: CHIS 2015 Methodology Overview ***6***

Chapter 2: CHIS 2015 Weighting and Imputation Plans ***16***

1. Overview **19**

1.1 CHIS 2015-2016 Sample Design 19

1.1.1 Telephone Numbers 19

1.1.2 Participating Adults 21

1.1.3 Participating Children and Adolescents 21

1.1.4 Follow-up of Nonrespondents 22

1.2 Quality Control 22

1.3 Remaining Section of this Plan 23

2. Response Status **25**

2.1 Final Response Status 25

2.2 Partially Completed Interviews 26

3. Household One-Year Weights **27**

3.1 Base Weights 27

3.2 Korean, Vietnamese, and Japanese Surname List Adjustment 28

3.3 Combining Sample across Period, Landline Frame, and Vendor 29

3.4 Phase 2 Subsampling Adjustment 30

3.5 Unknown Study Eligibility 32

3.6 Screener Nonresponse Adjustment 34

3.7 Multiplicity Adjustments 35

3.8 Final Weight 36

4. Adult One-Year Weights **37**

4.1 Base Weights 37

4.2 Phase 2 Adjustment 38

4.3 Adult Nonresponse Adjustment 38

4.4 Calibration Adjustment to NHIS 39
4.5 Composite Factor for Combining Cell Phone and LL Sample 40
4.6 Calibration Adjustment to Department of Finance Projections 41
4.7 Adult Analysis Weight 44

5. Child One-Year Weights **45**
5.1 Adjustment for Adult Nonresponse 45
5.2 Base Weights 46
5.3 Nonresponse Adjustment 47
5.4 Calibration Adjustment to NHIS 47
5.5 Composite Factor 47
5.6 Calibration Adjustment to Department of Finance Projections 48
5.7 Child Analysis Weight 48

6. Teen One-Year Weights **49**
6.1 Adjustment for Adult Nonresponse 49
6.2 Base Weights 50
6.3 Adjustment for Missing Parent Permission 50
6.4 Adjustment for Teen Nonresponse 50
6.5 Calibration Adjustment to NHIS 51
6.6 Composite Factor 51
6.7 Calibration Adjustment to Department of Finance Projections 51
6.8 Teen Analysis Weight 52

7. Imputation **53**

8. Variance Estimation **55**

References **56**

Endnotes **58**

Chapter 3: Additional Information on Weighting and Imputation ***61***

Dimensions used in adult weighting adjustments **62**

Variables and Calibration Levels for CHIS 2015 Adult Weights **63**

Item Nonresponse and Imputation Rates for Self-reported Sex and Age 67

Item Nonresponse and Imputed Rate for Self-reported Education 68

Pre- and Post-imputation Distribution Comparisons 69

Item Nonresponse for Self-reported Race and Ethnicity by Interview Type 72

Item Nonresponse for Phone Type Presence and Use 73

Listing of Variables used in Geocoding 74

Weighting Adjustment Magnitudes 75

TABLES

Number		Page
1.1	Design Strata: CHIS 2015-2016	20
1.2	Sampling Frame Used by Data Collection Quarter: CHIS 2015-2016	21
1.3	Comparison of CHIS Methodology by Study Year: CHIS 2015-2016	24
3.1	Final Response Status based on Phase-specific Information: CHIS 2015-2016	31
3.2	Candidate Variables for Screener Nonresponse Weight Adjustment: CHIS 2015-2016	35
4.1	Candidate Variables for Nonresponse Adjustment Applied to the Adult Weights: CHIS 2015-2016	39
4.2	Candidate Variables and Levels for Calibration Adjustment Applied to the Adult Weights: CHIS 2015-2016	42
7.1	Candidate Variables for Imputation: CHIS 2015-2016	54

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.

1.4 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

[1] http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201605.pdf

Tagalog. Tagalog interviews were conducted for part of the CHIS 2013-2014 cycle, but 2015 is the first year that Tagalog interviewers 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
Pre-paid	
Landline sample matched to address	$2
Promised	
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 continues 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 Weighting and Imputation Plans

The weighting plan used for CHIS 2015-2016 is included below for reference. Pages have been re-numbered to be consistent with this report, but tables have not been re-numbered. Additional tables showing results of the weighting are below in Chapter 3. A complete review of the weighting plan and results for the full cycle will be included in the full cycle reports.

February 2016

California Health Interview Survey 2015-2016
Weighting and Imputation 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

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

RTI Project Number 0214657.003.009

RTI Project Number
0214657.003.009

California Health Interview Survey 2015-2016
Weighting and Imputation Plan

February 5, 2016

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

1. Overview

Researchers apply analysis weights to survey responses to produce estimates for the target population with minimal biases and factors that could lower precision. This report provides an overview of the proposed weighting methodology for CHIS 2015-2016. We additionally discuss imputation methods planned for use to address item nonresponse in variables required for weighting.

In this first section of the plan, we provide a brief overview of the CHIS 2015-2016 sampling design (Section 1.1). Section 1.2 contains an overview of our quality control procedures used in constructing six sets of analysis weights—adult, child, and teen linearization weights plus a complementary set of replication weights. We conclude this overview in Section 1.3 with a summary of the remaining sections in this plan.

1.1 CHIS 2015-2016 Sample Design

The sample design for CHIS 2015-2016 is summarily described as a stratified two-stage dual-frame design (Phase 1) with a follow-up study on a subsample of nonrespondents (Phase 2). The strata are consistent with the 2013-2014 design and are shown in **Table 1.1**. At most, four CHIS questionnaires are completed: screener, adult, child (proxy) and teen. A visual representation of the study design is provided in ***CHIS 2015 DLV 9.1 Weighting Imputation Plan.2015_1230.vsd***.

Below, we briefly summarize relevant steps in the sampling process that inform the creation of the analysis weights. Additional details relevant for calculating the analysis weights are provided in the subsequent sections of the plan.

1.1.1 Telephone Numbers

Telephone numbers were randomly sampled on a quarterly basis from four landline list frames and one cell phone list frame. Specifically, RTI requested samples from the following frames:

- random-digit-dial (RDD) landline frame (LL),
- Vietnamese surname landline frame (SV),
- Korean surname landline frame (SK),
- Japanese surname landline frame (SJ), and
- RDD cellular telephone frame (CE).

Table 1.2 contains a listed of the frames used to date by CHIS data collection quarter.

Vendors tested all numbers for working status and submitted all cases regardless of status to RTI. RTI tested the landline samples again prior to release for data collection to remove non-working numbers. Information was gathered in the screener questionnaire to determine the eligibility of the household (landline) and the person answering the phone (landline and cell).

Table 1.1 Design Strata: CHIS 2015-2016

1 – Los Angeles (all)	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	

Table 1.2 Sampling Frame Used by Data Collection Quarter: CHIS 2015-2016

Quarter [2]	Sampling Frame [1]				
	LL	SV	SK	SJ	CE
2015 Q3	✓				✓
Q4	✓	✓	✓	✓	✓
2016 Q1	✓				✓
Q2	✓	✓	✓	✓	✓
Q3	✓	✓	✓	✓	✓
Q4	✓	✓	✓	✓	✓

Note: Q = quarter.

[1] RDD landline (LL), Vietnamese surname landline (SV), Korean surname landline (SK), Japanese surname landline (SJ), and RDD cellular (CE)

[2] The choice of frames for 2016 Q2-Q4 is speculative as of January 20, 2016.

1.1.2 Participating Adults

One person was selected to complete the adult interview using a different methodology for landline phone versus cell. For the landline numbers, the Rizzo method of selection was used to select one adult from a sampled household (Rizzo et al., 2004) for households with three or more adults. This method did 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. One adult was randomly chosen for the study with equal probability in two-adult households; the sole adult in a single-adult household was recruited for the study.

Conversely, for the cell sample, we assumed that the adult answering the sampled cell phone number was the sole user. As with single-person households in the landline sample, this person was recruited for the study.[2]

1.1.3 Participating Children and Adolescents

Once an adult was selected for the study, one child and one adolescent was randomly chosen among those under legal guardianship of the selected adult. Children ages 0-5 years were twice as likely to be selected as children ages 6-11; one adolescent (age 12-17) was chosen with equal probability among all eligible adolescents.

Details of the sampling methodology are as follows. For cell phone cases, the screener respondent and adult respondent were one in the same.[3] The number of children under 12

[2] Questions used previously to assess whether the cell phone was a shared device among multiple adults were excluded from the screener in CHIS 2015 in an effort to decrease the length of the screen and consequently increase participation.

[3] Note that cell phone cases discussed here include cell phone sample, ported landline numbers identified prior to fielding, and any landline telephone numbers identified as a cell phone during data collection.

years of age and teens 12-17 years of age was collected in the screener. Rostering and sampling children for a proxy interview and the adolescent occurred in Section G of the adult questionnaire. The probability of selecting a child in the 0-5 year age group was defined as $2n_{1i} / (2n_{1i} + n_{2i})$, where n_{1i} was the number of eligible children ages 0-5 years and n_{2i} was the number of children ages 6-11 years within household *i*. The corresponding selection probability for eligible children ages 6-11 years was $n_{2i} / (2n_{1i} + n_{2i})$. One adolescent was chosen with equal probability, i.e., the inverse of the number of adolescents in the household who are under the legal guardianship of the selected adult.

For landline sample cases linked to a one-adult household or where the screener and adult respondent were one in the same, the child/adolescent sampling methodology follows the steps specified for the cell phone cases above. Otherwise, if the screener respondent was not the chosen adult, then rostering and sampling occurred in the screener using the probabilities of selection noted above for children 0-5 years, children 6-11 years, and adolescents 12-17 years. The child-first method was employed if the selected adult was not available for to complete the interview but a parent of the selected child/adolescent was available to complete the child interview.

1.1.4 Follow-up of Nonrespondents

Sampled cases lacking the requisite number of interviews and telephone numbers without a known eligibility status from the first phase of the data collection (i.e., Phase-1 nonrespondents) were eligible for Phase-2 subsampling. Higher incentives than those stated for Phase 1 were promised to Phase 2 participants, to maximize response to Phase 2 in an attempt to reduce nonresponse bias in key estimates within the CHIS data.

1.2 Quality Control

RTI regularly implements a series of quality control procedures to ensure the accuracy of survey weights. For example, we compare the sum of the weights, an estimate of the target population, against external counts such as those tabulated from the American Community Survey. Large differences may indicate either errors or potential problems in model-based adjustments to these control totals.

We also compare statistics of the weights before and after an adjustment. Large differences before and after a weight adjustment is applied would indicate the need for further review. For example, a large relative change in an unequal weighting effect (UWE; i.e., design effect associated with the weights) calculated by certain important domains (e.g., race/ethnicity or geographic location) would be evaluated to determine if additional or new variables should be used for the weight adjustment model.

We will also examine the weights for outliers using the *3×interquartile range* rule overall and within key domains (see, e.g., Chen et al., 2014), in addition to an investigation of the

extreme weight methods employed for CHIS 2013-2014. Outliers are subject to trimming only after a thorough review of the weight components and only after discussions with UCLA (see endnote [1]).

Finally, we will compare five estimates from the current round of CHIS with CHIS 2014 as a preliminary trend evaluation. We will identify these five estimates in collaboration with UCLA (see endnote [2]).

All weight components and final analysis weights will be delivered to UCLA for subsequent review, along with a summary of the statistics for inclusion in the CHIS 2015-2016 methods reports (see endnote [3]). Revisions to the proposed plan will be discussed with UCLA prior to implementation.

1.3 Remaining Section of this Plan

The remaining sections of the plan detail our procedures for calculating CHIS 2015-2016 weights. The weighting process involves the creation of an initial weight for each sample unit and then a series of weight adjustments that are combined into a single final weight. This final weight is used for analyses. The weighting process begins with procedures to classify the response status of all sample cases released for data collection (Section 2). Next, we describe in Section 3 household-level weights generated as inputs to the three person-level weights. Steps to create weights for analyzing the adult data are described in Section 4, followed by the child (proxy) interview weight (Section 5) and the teen interview weight (Section 6). Having defined our weighting methodology, we discuss procedures for imputing missing values in variables required for weighting within Section 7. Note that the imputation task will commence prior to applying any nonresponse adjustments to the weights. In Section 8, we conclude this weighting and imputation plan with a description of two variance estimation techniques required for CHIS, along with the weights needed for each type. Throughout the document, we briefly comment on recommended changes from prior rounds of CHIS summarized below (**Table 1.3**).

Table 1.3 Comparison of CHIS Methodology by Study Year: CHIS 2015-2016

Chapter [1]	Section - Weight / Adjustment [1]	2013-14 [2]	2015 [1]	2016 [1]	Comments on comparison of 2015-16 vs. 2013-14
Household	1 - Base Weight	✓	✓	✓	
	2 - Surname List Adjustment		✓		
	3 - Combining Sample across Period, …	?	✓	✓	
	4 - Phase 2 Subsampling Adjustment		✓	✓	
	5 - Unknown Study Eligibility	✓	✓	✓	
	6 - Screener Nonresponse Adjustment	✓	✓	✓	Model-based approach recommended vs. weighting classes
	7 - Multiplicity Adjustments	✓	✓	✓	
	8 - Final Weight	✓	✓	✓	
Adult	1 - Base Weight	✓	✓	✓	
	2 - Phase 2 Adjustment		✓	✓	
	3 - Adult Nonresponse Adjustment	✓	✓	✓	Model-based approach recommended vs. weighting classes
	4 - Calibration Adjustment to NHIS	✓	✓	✓	Model-based approach will be investigated vs. poststratification
	5 - Composite Factor	✓	✓	✓	
	6 - Calibration Adjustment to Department of …	✓	✓	✓	Model-based approach will be investigated vs. poststratification
	7 - Adult Analysis Weight	✓	✓	✓	
Child / Teen	1 - Adjustment for Adult Nonresponse	✓	✓	✓	Slight variation on the input weights; model-based approach recommended vs. weighting classes
	2 - Base Weight	✓	✓	✓	Slight variation on within-HH selection probabilities
	3 - Nonresponse Adjustment	✓	✓	✓	Model-based approach recommended vs. weighting classes
	4 - Calibration Adjustment to NHIS	✓	✓	✓	Model-based approach will be investigated vs. poststratification
	5 - Composite Factor	✓	✓	✓	
	6 - Calibration Adjustment to Department of …	✓	✓	✓	Model-based approach will be investigated vs. poststratification
	7 - Child Analysis Weight	✓	✓	✓	

[1] See DLV 9.1 submitted on 12/30/2015.

[2] Based on review of methodology reports for CHIS 2011-12.

2. Response Status

Paramount to the creation of the weights is the classification of sample members into final response categories. Through this classification, we designate cases as being part of the target population (i.e., eligible), not part of the target population (i.e., ineligible), or eligibility status unknown. We discuss the mapping of data-collection status codes into four response status groups within Section 2.1. In Section 2.2, we briefly discuss the response status for partially completed interviews. We conclude with a brief discussion of weighted response rates in Section 2.3.

2.1 Final Response Status

RTI will classify sample cases into four response categories prior to starting the weighting process (AAPOR 2015):

- *Eligible respondent* = person sampled through eligible phone number and completes interview;
- *Eligible nonrespondent* = person sampled through eligible phone number and either refuses to participate or does not answer a sufficient number of items in the questionnaire (i.e., not a usable for analyses);
- *Ineligible sample member* = business, group quarters, non-working, household without at least one adult (18 years of age), cell phone link only to person(s) less than 18 years of age, other non-residential; and
- *Unknown eligibility* = all other cases

Table 2.1 in the accompanying Excel file contains the current mapping of final status codes to the four response categories.[4] We will review all sampled cases to determine whether their eligibility status is known. For example, all refusals without a defined eligibility status will be reclassified as unknown eligibility and cell phone non-contact cases will be evaluated for an ineligible "number active but not assigned to a person" status (e.g., no voicemail set-up). Additionally, sampled cases designated as partial completes will be reassigned prior to weighting as either eligible respondents or eligible nonrespondents based on the completeness of the interview (see Section 2.2).

[4] Note that this mapping undergoes a periodic review throughout data collection to ensure quality and consistency with past rounds of CHIS. This list will finalized and approved by UCLA prior to implementation of the weighting methodology.

2.2 Partially Completed Interviews

Partial interviews are those with sufficient information for analyses obtained from an eligible respondent. The definition of sufficiency is an interview completed at least through the end of Section K. We will classify incomplete interviews (adult, child, and teen) with less information as eligible nonrespondents.

3. Household One-Year Weights

The first stage of selection for CHIS 2015-2016 is the household by way of either a sampled landline telephone number or a sampled cell phone number. Note that the weights generated at this stage are called "household annual weights" to keep with the historic CHIS label. These weights by themselves should not be used to generate estimates for the household population in California because they do not incorporate important adjustment factors related to nonresponse within the household.

In this section, we detail the proposed plan for calculating a household-level analysis weight by sampling frame. We will use this finalized weight as the basis for the instrument-specific analysis weights—adult, child (proxy), and teen—discussed in the subsequent sections of this plan.

Specifically, we define the initial base weights by sampling frame (Section 3.1), followed by an adjustment applied to the base weights of cases sampled from the surname list (Section 3.2). A unifying adjustment is proposed in Section 3.3 to combine across periods of data collection within a single year of the study, across four landline list frames, and for CHIS 2015 only, across sample vendor. We define a subsampling adjustment for the second phase of the design (Section 3.4). Weights for sampled cases with a known eligibility status are adjusted for those without a known eligibility status (Section 3.5); weights for the unknown eligibility status cases are then set to zero. Next, we apply an adjustment for household-level nonresponse defined as sampled cases without a completed screener (Section 3.6). We include at least one multiplicity factor for eligible participating households to address multiple paths of selection into the sample (Section 3.7). The final household weight is defined in Section 3.8 after applying any necessary refinements. As discussed in Section 1.2 of this report, quality assessment and control procedures are administered throughout the process (see Section 1.2 for the quality control procedures).

3.1 Base Weights

A base weight, also referred to as a "design weight" or "sampling weight", adjusts only for the specific process of sampling from the sampling frame. For the household weights, phone number is our proxy for household; the base weight is calculated as the inverse of the probability of selection for each sampled phone number from the respective frame of all phone numbers. This weight serves as the basis for the three CHIS analysis weights and the associated replicate weights used for variance estimation (see Section 8 of this plan)The base weights are calculated separately for each sampling frame from which the phone numbers were selected (e.g., landline RDD frame, cell phone RDD frame).

To date, CHIS 2015-2016 includes quarterly samples from each of five types of list frames—four landline frames and one cell phone frame. Specifically, RTI requested samples from the following frames: RDD landline (LL), Vietnamese surname landline (SV), Korean surname landline (SK), Japanese surname landline (SJ), and RDD cellular (CE). For CHIS 2015, RTI purchased LL and CE samples from two vendors—Marketing Systems Group (MSG) and Survey Sampling International (SSI)—for comparative purposes only. We purchased surname frame samples (SV, SK and SJ) from MSG only.

The base weights will be calculated as follows in keeping with the stratified simple random sampling design within sampling frame.

$$HW0_{VFQhi} = \frac{N_{VFQh}}{n_{VFQh}} \tag{3.1}$$

where V indexes the vendor (MSG and SSI); F indexes the sampling frame (LL, SV, SK, SJ, and CE); Q indexes the period of data collection (i.e., quarter); h indexes the design strata; i indexes the household; n_{VFQhi} is the size of the sample selected within stratum h from vendor V, sampling frame F, and quarter Q; and N_{VFQhi} is the associated frame count. Note that index V is not applicable for CHIS 2016 because we will purchase samples only from MSG.

3.2 Korean, Vietnamese, and Japanese Surname List Adjustment

The MSG surname list frames were a proper subset of the RDD landline frame (i.e., they only included landline numbers as known by the vendor). Ideally, landline samples for CHIS 2015 would have been drawn from four mutually exclusive frames. However, because surname lists were not used initially as part of the CHIS 2015 design, LL was not constructed in a way to exclude the surname frame phone numbers to achieve the desired exclusivity for the Quarter 4 samples. Thus, the surname list sample could have been chosen either from the respective surname frame or from LL. To address the fact that surname samples had twice the chance of being selected for CHIS 2015, the following adjusted base weight will be created:

$$HW1_{VFQhi} = \begin{cases} HW0_{VFQhi}, & F=\text{LL and CE} \\ HW0_{VFQhi} \times 0.5, & F=\text{SV, SK, and SJ (MSG only)} \end{cases} \tag{3.2}$$

Note this adjustment will not be applied to the CHIS 2016 data because we plan to draw sample cases from the four mutually exclusive landline frames.

A similar adjustment should be applied to the LL sample cases that could have been selected from the surname list; however, this adjustment is problematic. Phone numbers are included on the surname landline frame only if the number is listed in the white pages *and* the last name has some defined likelihood of being the nationality in question. Phone numbers are known to be listed or unlisted based on whether the phone number had a matched address (a process conducted by MSG). The association between surname and the surname frame is not known because the likelihood "model" is proprietary to MSG.

RTI will evaluate the need to send the responding LL cases to MSG indicating either a Vietnamese, Korean, or Japanese nationality. Those listed on the surname frames would have the adjustment in (3.2) applied to their base weights. Currently, however, the number of cases from the surname frames is small and the adjustment would have limited impact on the weights, suggesting that such a step may not be cost effective.

3.3 Combining Sample across Period, Landline Frame, and Vendor

Multiple independent samples were selected for each year of the 2015 and 2016 CHIS survey years. We selected independent samples for each quarter of data collection to ensure the sampling frame information was current. As noted in Section 3.2, RTI also selected landline telephone numbers from four MSG sampling frames—LL, SV, SK and SJ. Finally, within the two quarters of CHIS 2015, RTI conducted an experiment to quantify differences between two sample vendors—MSG and SSI—that could potentially benefit future rounds of CHIS.

The data from the independent LL and CE samples will be combined through a poststratification adjustment. Since the surname landline list frames were completely covered by the LL frame, we will additionally combine the MSG surname samples (selected for only CHIS 2015 Quarter 4) with the LL cases; as noted previously, we will not need this adjustment for the 2016 design. The poststratification adjustment will be created within the design strata and have the following form:

$$HW2_{Fhi} = \begin{cases} HW1_{VQFhi} \times \dfrac{N_{LLh\bullet}}{\sum_V \sum_Q \sum_F \sum_{i \in s_h} HHW1_{VQFhi}}, & F = \text{LL, SV, SK, and SJ} \\ HW1_{VQFhi} \times \dfrac{N_{CEh\bullet}}{\sum_V \sum_Q \sum_{i \in s_h} HHW1_{VQFhi}}, & F = \text{CE} \end{cases} \quad (3.3)$$

where V indicates samples selected by each vendor (V=MSG and SSI); Q indicates samples selected for each the period of data collection (i.e., Quarters 3 and 4 for CHIS 2015 and Quarters 1-4 for CHIS 2016); F is the original sampling frame; and s_h is the sample within design stratum h.

Minimal differences were found in our comparisons of the two vendors with the CHIS 2015 Quarter 3 data. See Deliverable 1.3 Quarter 3 sent to UCLA on 11/25/2015. RTI will proceed in 2016 with MSG as the vendor. Since we will continue purchasing quarterly samples, this poststratification adjustment is appropriate for CHIS 2016. We propose to use the most current Quarter 4 stratum counts from MSG, both landline ($N_{LLh\bullet}$) and cell ($N_{CEh\bullet}$), as the poststratification totals in the CHIS 2015 adjustment to maintain consistency with the vendor used in CHIS 2016.

After combining across the four landline frames, we will now refer to the combined LL, SV, SK, and SJ samples as the landline-surname sample (LS). Verifying that the sum of the weights in (3.3) by frame type (landline, cell phone) frame and stratum (h) equals the poststratification totals, $N_{LLh\bullet}$ and $N_{CEh\bullet}$, is an example of a quality check that will be administered throughout the weighting process.

3.4 Phase 2 Subsampling Adjustment

A new responsive design component was introduced with CHIS 2015. Referred to as a nonresponse follow-up (NRFU), or the second phase of a two-phase design, a random subsample of known (or possibly) eligible cases who have yet to respond in the first phase were re-contacted using a modified recruitment protocol. Consequently, we will establish the response status for the sample cases within each phase of the design.

Following the information presented in Section 2.1, we designate a final response status based on the combination of Phase 1 and Phase 2 results (**Table 3.1**).

Table 3.1 Final Response Status based on Phase-specific Information: CHIS 2015-2016

Response Status		
Final[1]	**Phase 1**	**Phase 2**[1,2]
Eligible Respondent	Eligible Respondent	*na*
	Eligible Nonrespondent	Eligible Respondent
	Unknown	Eligible Respondent
Eligible Nonrespondent	Eligible Nonrespondent	Eligible Nonrespondent
	Unknown	Eligible Nonrespondent
Ineligible	Ineligible	*na*
	Unknown	Ineligible
Unknown	Unknown	Unknown
not sampled	Eligible Nonrespondent	*not sampled*
	Unknown	*not sampled*

[1] *not sampled* = cases not sampled for Phase 2 and removed from the file after applying a weight adjustment
[2] *na* = completed Phase-1 case that was not eligible for Phase 2

Additionally, we classify cases for this adjustment as either a landline or a cell phone number based on the sampling frame, pre-screening information, or initial responses to the screener. For convenience, we refer to the combined LL, SV, SK, and SJ samples as the augmented landline sample (LS) as before, and the combined cell phone and landline numbers ported to a cell phone as the cell/ported sample (CP) indexed by *T*. Note that all sampled landline numbers ported to a cell phone will be analyzed as cell phone numbers from this point forward.

A stratified Phase-2 sample was selected from the LS and CP Phase-1 samples from four strata:

- Screener Not Complete, No Contact
- Screener Not Complete, Some Contact
- Screener Complete, Adult-Only Household
- Screener Complete, Child/Teen Household

Design stratum was used as a sorting variable prior to drawing the sample to ensure geographic representation across the California counties.

The following weight adjustment reflects the Phase-2 subsampling:

$$AH3_{Thki} = \begin{cases} 1, & \text{Phase-1 respondent or study ineligible} \\ N_{Thk}/n_{Thk}, & \text{Phase 2 frame, sampled} \\ 0, & \text{Phase 2 frame, not sampled} \end{cases} \quad (3.4)$$

where T indicates the phone type (LS, CP); k (k=1–4) indicates the Phase-2 stratum; N_{Thk} is the number of sample cases eligible for Phase-2 sampling by sample type and strata; and n_{Thk} is the number selected. Note that $AH3_{Thk}$=1 indicates cases not eligible for Phase-2 selection. Note that we will adjust n_{Thk} and N_{Thk} to reflect any cases either randomly chosen and not released for data collection, or released for data collection but never dialed. This adjustment is then applied to the weight in (3.4) to form:

$$HW3_{Thi} = HW2_{Fhi} \times AH3_{Thki} \quad (3.5)$$

As shown in (3.4), cases eligible but not sampled for Phase 2 are effectively removed from the weighting process at this stage since their adjustment $AH3_{Thk}$ is zero. The weights for the Phase-2 sample cases are inflated to account for the subsampling, much in the vein of Phase 1 cases having a (positive) weight to account for all those not included in the original CHIS sample.

3.5 Unknown Study Eligibility

Telephone numbers are designated as ineligible for CHIS if they meet any of the following criteria. We categorize ineligibility as follows:

- Non-working numbers
- Non-residential (e.g., business/government, fax/modem, active cell numbers not assigned)
- Ineligible residence (e.g., all residents/cell phone owner less than 18 years of age, institutionalized residence, group quarters, 9 or more unrelated persons)

Multiple attempts were made to obtain information to determine the eligibility status of the sample in both phases of the design. For example, RTI evaluated landline numbers prior to data collection and withheld any found to be non-working (see endnote [4]). Other ineligible sample cases were identified during data collection.[5] After approval from UCLA, we plan to assign (logically impute) a few sample cases to an ineligible status based on

[5] RTI made a minimum of 10 call attempts to all cases before finalizing cases as eligible, not eligible, or eligibility unknown. These call attempts spanned morning, midday, and evening hours as well as weekday and weekend periods.

paradata. For example, RTI will reassign "non-contact" cell phone cases without any indication of a working voicemail system to a "IW - Nonworking #" status.

There remains, however, a set of cases without a known (or imputed) eligibility status. We follow the lead documented for CHIS 2011-2012 and distinguish by the type of ineligibility noted above. Weights for the cases with unknown residential status will be distributed across the sample cases with known residential status using a weighting class adjustment similar in form to (3.3). Namely,

$$HW4_{Thi} = \begin{cases} HW3_{Thi} \times \left(\dfrac{\sum_{i \in s_g} HW3_{Thi}}{\sum_{i \in s_g} \left[I_i\left(status\ known\right) \times HW3_{Thi} \right]} \right), \\ \qquad \text{Phones with known residential status} \\ 0, \qquad \text{Phones with unknown residential status} \end{cases} \tag{3.6}$$

for T=LS and CP, where $I_i(status\ known)$=1 if the residential status for the sampled phone number i is known (eligible or ineligible), and $I_i(status\ known)$=0 otherwise within a total of G weighting classes (g=1, …, G).[6]

A second adjustment of similar form will address cases known (imputed) to be residential but the within-residence eligibility status is not fully known. The associated weight takes the following form:

$$HW5_{Thi} = \begin{cases} HW4_{Thi} \times \left(\dfrac{\sum_{i \in s_g} HW4_{Thi}}{\sum_{i \in s_g} \left[I_i\left(HHstatus\right) \times HW4_{Thi} \right]} \right), \\ \qquad \text{Residence with known eligibility status} \\ 0, \qquad \text{Residence with unknown eligibility status} \\ HW4_{Thi}, \qquad \text{Non-residential phone} \end{cases} \tag{3.7}$$

where $I_i(HHstatus)$=1 if the within-household eligibility status for the residence is known (eligible or ineligible), and $I_i(HHstatus)$=0 otherwise. Note that only ineligible sample cases that are residences are included in this adjustment; non-residential phone numbers are excluded from the adjustment.

The weighting classes, denoted as g in (3.6) and (3.7), will be defined with variables available for *all* sample cases. Currently, we propose to use only sample type (T=LS, CP)

[6] The number of weighting classes may differ by sample type (LS and CP); we use a single index G only for convenience.

and design stratum (*h*); we plan to evaluate the predictive power of additional variables (e.g., area code) with the known eligibility cases, keeping in mind that weighting classes with sufficient sample sizes are desired (see, e.g., Valliant et al., 2013).

3.6 Screener Nonresponse Adjustment

The next adjustment in the series to finalize the household-level weight will address CHIS screener nonresponse among those cases known or imputed to be eligible. In other words, sample cases eligible but not sampled for Phase 2, those known to be ineligible, and those with an unknown eligibility status are removed from the weighting process at this point.

Past publically available CHIS methodology reports indicate that a weighting class adjustment, much like those discussed previously, was used to account for screener nonresponse. Weighting classes (i.e., groups) are formed by combining binary, categorical, or categorized continuous variables thought to be associated with response and preferably also with characteristics of importance from the study. As noted in Kim et al., (2007), use of too many variables can result in too many or even small (empty) weighting classes that hinder the calculation of an efficient nonresponse-adjusted weight. Determining an effective mechanism for collapsing small cells can be a time-consuming process, yielding minimal gains in precision (via reduced variations in weights) and possibly limiting the reduction of bias attributable to nonresponse. Consequently, incorporating only a few variables limits the capacity to reduce nonresponse bias, the true goal of this weight adjustment. Therefore, RTI will investigate a more refined model-based methodology, using a model within the SUDAAN® WTADJUST procedure (RTI, 2012).

PROC WTADJUST enables the creation of a model-based adjustment that can be used either directly or to form weighting classes by categorizing the response propensities into at least five groups. Additionally, for the direct method, the procedure enables adjustments to be constrained within a defined set of values to maintain control over the variation of the weights. Both approaches by sample type (LS, CP) will be evaluated by comparing calculated unequal weighting effects and measures of nonresponse bias reduction. Additionally, we will evaluate the need for separate models to address nonresponse for screeners with varying levels of item response. See endnotes [5] and [6].

Candidate variables for the nonresponse adjustment model, known for respondents and nonrespondents, are shown in **Table 3.2**. Any variables with insufficient sample size across the various levels will be eliminated. We will determine the final set of model covariates and any relevant interactions using regression trees (CART; Breiman et al., 1984) to develop a parsimonious model (see endnote [7]). Note that the analyses may suggest the

use of one model for nonrespondents with limited/no information obtained from the screener, and another model for nonrespondents with some screener information obtained from the screener.

Table 3.2 Candidate Variables for Screener Nonresponse Weight Adjustment: CHIS 2015-2016

Level	Variable Description
Geographic/ Sampling	• Design stratum / county • Area code • Phase 2 sample member
Household	• Phone usage (home only or home/business) • Is anyone in your household covered by Medi-CAL? • Number of adults (*possibly categorized*) • Number of children 0-11 years old (*possibly dichotomized/categorized or collapsed into variable for children 0-17 years old*) • Number of teens (*possibly dichotomized/categorized or collapsed into variable for children 0-17 years old*) • Presence of spouse in household • Child-first household • Multiple phones per household
Person	• Does selected adult have any medical conditions?

* Information must be available for respondents and nonrespondents to be used in the adjustment model.

The resulting nonresponse-adjusted weight is defined as

$$HW6_{Thi} = \begin{cases} HW5_{Thi} \times AH6_{Thi} & \text{Screener respondent} \\ 0, & \text{Screener nonrespondent} \end{cases} \quad (3.8)$$

where $HW5_{Thi}$ is defined in (3.7) and $AH6_{Thi}$ is the model-based nonresponse adjustment with final form to be determined.

3.7 Multiplicity Adjustments

Two adjustments are proposed to address multiple ways in which a sampled household could have been selected for CHIS. The first adjustment addresses multiple telephone numbers within the same sampling frame. RTI proposes to adopt the adjustment used in past rounds of CHIS. Namely,

$$AH7_{Thi} = \begin{cases} 0.5, & T = \text{LS with multiple residential numbers} \\ 1, & \text{otherwise} \end{cases} \quad (3.9)$$

resulting in an adjusted weight of the form

$$HW7_{Thi} = HW6_{Thi} \times AH7_{Thi} \tag{3.10}$$

If needed, a second adjustment will be applied to adjust for households actually selected through multiple telephone numbers. The resulting weight will take the form:

$$HW8_{Thi} = \begin{cases} HW7_{Thi} \times \left(\dfrac{\sum_{i'} HW7_{Thi'}}{HW7_{Thi}} \right), & \text{Household with duplicate number, first interview} \\ 0, & \text{Household with duplicate number, not interviewed or subsequent interview} \\ HW7_{Thi}, & \text{Unique phone numbers} \end{cases} \tag{3.11}$$

3.8 Final Weight

We will make any final modifications to the adjusted weight, $HW8_{Thi}$ in (3.11), to create the final weight for household *i* in design stratum *h* with sample type T (LS, CP),

$$HHW_{Thi} = HW8_{Thi} \times AH9_{Thi} \tag{3.12}$$

where $AH9_{Thi}$ is the final adjustment(s) with final form to be determined. Adjustments may include calibration to external population counts or if needed, a trimming factor to reduce the variation in the weights.

4. Adult One-Year Weights

The second stage of selection for CHIS 2015-2016 is person(s) within household, depending on the composition of the household. Below, we detail the proposed plan for calculating an adult analysis weight for analyzing the annual CHIS data files. Specifically, we define the initial base weights for the randomly selected adult in Section 4.1. We discuss the adjustment applied to the base weights for those completing the interview in the first design phase within Section 4.2. Nonresponse to the adult interview request is addressed next (Section 4.3), followed by a calibration adjustment to population control totals in Section (4.4). A composite factor is introduced in Section 4.5 to combine dual users (landline and cell) selected from either the landline or cell sampling frames. We conclude the proposed weighting steps in Section 4.6 with a final calibration adjustment, but note that further discussion is needed to address any challenges experienced with CHIS weights from prior rounds. Any additional adjustments will result in the final adult analysis weight (Section 4.7). RTI will administer quality assessment and control procedures throughout the process as discussed in Section 1.

4.1 Base Weights

One eligible adult was chosen from the sampled household. For landline telephone numbers, one adult was selected with equal probability from all those residing in the household. The adult answering the sampled cell phone was assumed to be the sole owner and therefore selected with probability one. Additional information on the sample selection procedure is provided in Section 1. As a result, the j^{th} adult base weight is defined as

$$AW1_{Thij} = HHW_{Thij} \times AA1_{Thij}, \tag{4.1}$$

a function of the final household weight, HHW_{Thi} given in (3.12), and the within-household adult selection weight calculated as follows:

$$AA1_{Thij} = \begin{cases} ACNT_{Thij}, & T = \text{LS} \\ 1, & T = \text{CP} \end{cases} \tag{4.2}$$

Prior to the calculation, RTI will evaluate the variation in $ACNT_{Thij}$, the total number adults in the household. If using the actual number in the weight adjustment introduces a large increase in the UWE, we will consult with UCLA about setting a maximum value, much like the adjustment for number of telephone lines in (3.9).

4.2 Phase 2 Adjustment

Section 3.4 detailed the adjustment applied to the household-level weight to account for the multiphase CHIS design introduced in 2015. Households lacking the requisite number of interviews (e.g., three interviews in a household with both an eligible child and teen) and telephone numbers without a known eligibility status were eligible for Phase 2. To account for adult interviews in Phase-2 households that were actually completed in Phase 1, we will apply a factor to remove the Phase-2 subsampling from the base weights in (4.1). Namely,

$$AA2_{Thik} = \begin{cases} n_{Thk}/N_{Thk}, & \text{Selected for Phase 2 but adult interview completed in Phase 1} \\ 1, & \text{Otherwise} \end{cases} \quad (4.3)$$

where T indicates the sample type (LS, CP); h the design stratum; and k the Phase-2 stratum; N_{Thk} is the number of sample cases eligible for Phase-2 sampling; and n_{Thk} is the number selected for Phase 2. Note that (4.3) is the inverse of the adjustment applied to $HW3_{Th}$, i.e., $AH3_{Thk}$ defined in (3.4). The resulting adjusted weight will have the form,

$$AW2_{Thij} = AW1_{Thij} \times AA2_{Thk} \quad (4.4)$$

4.3 Adult Nonresponse Adjustment

As with the CHIS screener, we made repeated attempts to obtain a completed adult questionnaire. However, at least one nonresponse adjustment will be made to the weights to address any nonresponse bias. The nonresponse-adjusted weight will take the form:

$$AW3_{Thij} = \begin{cases} AW2_{Thij} \times AA3_{Thij} & T=\text{LS and CP respondents} \\ 0, & T=\text{LS and CP nonrespondents} \end{cases} \quad (4.5)$$

where $AW2_{Thij}$ is defined in (4.4). The multiplier $AA3_{Thij}$ is the nonresponse adjustment derived using much of the same procedures as discussed for the screener nonresponse-adjusted weight (Section 3.6). Specifically, we will use correlational analysis and statistical methods such as CART to determine a parsimonious set of variables for use in a model-based nonresponse adjustment. We will again use SUDAAN's PROC WTADJUST to generate the response propensities; these values will be used directly (or indirectly as with a

weighting class adjustment) to form $AW2_{Thij}$. Candidate variables for the nonresponse model(s) are provided in **Table 4.1**.

Table 4.1 Candidate Variables for Nonresponse Adjustment Applied to the Adult Weights: CHIS 2015-2016

Level	Description
Geographic/ Sampling	• Design stratum / county • Area code • Phase 2 sample member
Household	• Phone usage (home only or home/business) • Is anyone in your household covered by Medi-CAL? • Number of adults (*possibly categorized*) • Number of children 0-11 years old (*possibly dichotomized/categorized or collapsed into variable for children 0-17 years old*) • Number of teens (*possibly dichotomized/categorized or collapsed into variable for children 0-17 years old*) • Presence of spouse in household • Child-first household • Multiple phones per household
Person	• Does selected adult have any medical conditions? • Sampled adult was screener respondent • Sex of adult respondent • Age of adult respondent (continuous or categorized [18-30 years old [yo], 31-45 yo, 46-65 yo, and 65+]) • Phone use status (landline only; cell only; dual-use landline sample; dual-use cell sample)

* Information must be available for respondents and nonrespondents to be used in the adjustment model.

RTI will evaluate the data for differential patterns of nonresponse across groups such as phase sample, sample type, and race/ethnicity to determine if separate adjustments would be advantageous to reduce nonresponse bias. For example, child-first households in certain regions of the state may have different participation rates, thus suggesting the use of different models to address nonresponse for child-first and other households.

4.4 Calibration Adjustment to NHIS

The overall sum of the nonresponse-adjusted weights, $AW3_{Thij}$ in (4.5), by phone-use status will estimate the total adults in the population of landline users, cell phone users, and dual users in California. However, misalignment of the estimates with reality may occur since the three mutually exclusive frames, one for each phone-use status, does not exist to draw our samples. Thus, a calibration adjustment is proposed to align the weight sums for the four groups generated with the CHIS data: landline only (L), cell phone only (C), dual-use landline sample (DL), and dual-use cell sample (DC).

As in past rounds, we propose to use estimates from the National Health Interview Survey (NHIS) as the source for the population control totals (Blumberg & Luke, 2015). RTI will include the overall population counts by phone-usage status (landline only, cell only, dual use) in the calibration model; we will also investigate the utility of more refined counts by

demographic group such as sex and race/ethnicity. The sum of the calibrated weights in (4.6) will be verified against the control totals used in the adjustment.

We will calculate the calibrated weight as:

$$AW4_{Thij} = \begin{cases} AW3_{Thij} \times AA4_{Lhij} & \text{landline-only users} \\ AW3_{Thij} \times AA4_{DLhij} & \text{dual-use landline sample} \\ AW3_{Thij} \times AA4_{DChij} & \text{dual-use cell sample} \\ AW3_{Thij} \times AA4_{Chij} & \text{cell phone-only users} \end{cases} \quad (4.6)$$

where the model-based adjustment factors for landline-only users ($AA4_{Lhij}$), cell phone-only users ($AA4_{Chij}$), and dual users ($AA4_{DLhij}$ and $AA4_{DChij}$) will be calculated with SUDAAN's PROC WTADJUST using NHIS control totals. Note that WTADJUST enables the calculation of adjustment for either nonresponse or calibration (see, e.g., Kott, 2006).

4.5 Composite Factor for Combining Cell Phone and LL Sample

The calibrated survey weights for our CHIS respondents now align with population estimates (controls) generated from a large national survey for two of the four CHIS groups—landline only and cell only. Summing the weights for the dual users, however, will estimate twice the size of the dual-use population. The next step is to adjust (composite) the weights for the dual-use cases so that combined they will reproduce the population size regardless of their frame source.

The composite weight will be constructed as:

$$AW5_{hij} = \begin{cases} AW4_{Thij} & \text{landline-only users} \\ AW4_{Thij} \times \lambda_A & \text{dual users, landline frame} \\ AW4_{Thij} \times (1 - \lambda_A) & \text{dual users, cell frame} \\ AW4_{Thij} & \text{cell phone-only users} \end{cases} \quad (4.7)$$

where $0 \le \lambda_A \le 1$ is the composite factor for the adult dual-use respondents. RTI will investigate various methods to determine λ_A such as setting the factor equal to

- 0.5, giving equal weight to both samples;
- a value that maximizes the precision for a set of important estimates identified in collaboration with UCLA;
- a value that minimizes bias for a set of important estimates (see, e.g., Brick et al., 2011).

We will allow compositing factors to vary across important analytic domains only if such differences do not decrease precision or increase bias for other key objectives. The sum of the calibrated weights in (4.7) will be verified against the control totals used in the original adjustment in (4.6).

4.6 Calibration Adjustment to Department of Finance Projections

In keeping with the specified statement of work for CHIS 2015-2016, RTI will calibrate the composite weights, $AW5_{hij}$ in (4.7), to the eligible population projections supplied by the State of California's Department of Finance. Population estimates associated with CA residents living in group quarters (e.g., nursing homes, prisons) and others who are not eligible for CHIS will be excluded from the population controls, using techniques similar to those documented in the CHIS 2011-2013 methods report (see section 7 of CHIS, 2014). Additional adjustments to the control totals will be applied as appropriate to maintain consistency with past rounds of CHIS (see endnote [8]). We will calculate the calibrated weight,

$$AW6_{hij} = AW5_{hij} \times AA6_{hij}, \tag{4.8}$$

as a function of the composite weight and a calibration adjustment factor $AA6_{hij}$ obtained from SUDAAN PROC WTADJUST. Candidate model covariates and interactions will mimic those used for raking in prior rounds of CHIS (**Table 4.2**). Additional variables may be identified and included in consultation with UCLA. The final list of variables and interactions (i.e., dimensions) will be determined based on factors such as CHIS respondent sample size, the availability (and precision) of calibration control totals, and the effect on bias and precision of the CHIS estimates.

We will verify that the sum of the calibrated weights reproduces the projected estimates before proceeding with the final step.

Table 4.2 Candidate Variables and Levels for Calibration Adjustment Applied to the Adult Weights: CHIS 2015-2016

Group	Characteristic	Levels[1]	Description
Geography	Reported Stratum (44)[2]	1	Los Angeles
		2	San Diego
		3	Orange
		4	Santa Clara
		5	San Bernardino
		6	Riverside
		7	Alameda
		8	Sacramento
		9	Contra Costa
		10	Fresno
		11	San Francisco
		12	Ventura
		13	San Mateo
		14	Kern
		15	San Joaquin
		16	Sonoma
		17	Stanislaus
		18	Santa Barbara
		19	Solano
		20	Tulare
		21	Santa Cruz
		22	Marin
		23	San Luis Obispo
		24	Placer
		25	Merced
		26	Butte
		27	Shasta
		28	Yolo
		29	El Dorado
		30	Imperial
		31	Napa
		32	Kings
		33	Madera
		34	Monterey
		35	Humboldt
		36	Nevada
		37	Mendocino
		38	Sutter
		39	Yuba
		40	Lake
		41	San Benito
		42	Tehama-Glenn-Colusa
		43	Del Norte-Siskiyou-Lassen-etc.
		44	Tuolumne-Calaveras-Amador-Inyo-etc.

(*continued*)

Table 4.2 Candidate Variables and Levels for Calibration Adjustment Applied to the Adult Weights: CHIS 2015-2016 (*continued*)

Group	Characteristic	Levels[1]	Description
Geography	Reported Los Angeles SPA [2]	1	Antelope Valley
		2	San Fernando Valley
		3	San Gabriel Valley
		4	Metro
		5	West
		6	South
		7	East
		8	South Bay
	Reported San Diego HSR [2]	1	North Coastal
		2	North Central
		3	Central
		4	South
		5	East
		6	North Inland
	Sub-state Region [3]	1	Northern & Sierra Counties (Butte, Shasta, Humboldt, Lake, Mendocino, Yuba, Nevada, Sutter, Colusa, Glenn, Tehama, Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity, Alpine, Amador, Calaveras, Inyo, Mariposa, Mono. Tuolumne)
		2	Greater Bay Area (Santa Clara, Alameda, Contra Costa, San Francisco, San Mateo, Sonoma, Solano, Marin, Napa)
		3	Sacramento Area (Sacramento, Placer, Yolo, El Dorado)
		4	San Joaquin Valley (Fresno, Kern, San Joaquin, Stanislaus, Tulare, Merced, Kings, Madera)
		5	Central Coast (Ventura, Santa Barbara, Santa Cruz, San Luis Obispo, Monterey, San Benito)
		6	Los Angeles
		7	Other Southern California (San Diego, Orange, San Bernardino, Riverside, Imperial)
	State excluding LA, SD [3]	1	Remainder of CA
Household	Number of adults in household	*various groups*	e.g., 1 adult, 2 adults, etc.
	Number of children in household	*various groups*	e.g., 0 children, 1 child, etc.
	Number of teens in household	*various groups*	e.g., 0 teens, 1 teen, etc.
Person	Person type	1	Adult
		2	Child
		3	Teen

(*continued*)

Table 4.2 Candidate Variables and Levels for Calibration Adjustment Applied to the Adult Weights: CHIS 2015-2016 (*continued*)

Group	Characteristic	Levels[1]	Description
Person	Age (categorized)	*various groups*	e.g., 0-5 years, 6-11 years, etc.
	Sex	1	Male
		2	Female
	Adult education	1	Less than High School
		2	High School grad or GED
		3	At least some college
	Race/ethnicity	1	Latino 18 years old or older
		2	Non-Latino White 18 years old or older
		3	Non-Latino African American 18 years+
		4	Non-Latino American Indian 18 years+
		5	Non-Latino Asian 18 years old or older
		6	Non-Latino Native Hawaiian 18 years+
		7	Non-Latino 2+ races 18 years+
	Asian nationality[4]	1	Non-Latino Chinese
		2	Non-Latino Korean
		3	Non-Latino Filipino
		4	Non-Latino Vietnamese
		5	Non-Latino Other Asian

[1] Levels may change when interacted with other control total variables to accommodate respondent sample size and available control totals. We will include the final list of dimensions in DLV 9.2 Weighting Documentation and in the Methods report (Task 10).

[2] Reported stratum (county) information.

[3] Variable derived from Reported stratum (county) information.

[4] RTI will investigate the utility of including additional Asian nationalities in the calibration adjustment.

4.7 Adult Analysis Weight

As with the household-level weight (Section 3.8), we will make any final adjustments to the weights based on a multistep quality review. Designating the adjustments as $AA7_{hij}$, the final adult analysis weight will be defined as

$$ADW_{hij} = AW6_{hij} \times AA7_{hij} \quad (4.9)$$

RTI will discuss additional weighting refinements with UCLA based on challenges experienced in prior rounds of the study before finalizing the weights.

5. Child One-Year Weights

Children, ages 11 years and younger, of the randomly chosen adult in the CHIS participating households are also eligible for the study. We collect information either from the adult participant or from the other legal parent who completed the screener on the selected child and not the child directly.

Below, we describe our proposed plan for calculating a child (proxy interview) analysis weight for analyzing an annual CHIS data file. Specifically, we define the input values for the child weights in Section 5.1 that are then adjusted to account for the child-level sampling (Section 5.2). We briefly describe one or more nonresponse adjustments applied to the weights in Section 5.3, followed by a calibration adjustment to population control totals in Section (5.4). A composite factor is mentioned in Section 5.5 to combine dual users (landline and cell) selected from the landline or the cell sampling frames similar to the method discussed for the adult weight. We propose a final calibration adjustment in Section 5.6, but leave room for additional refinements to create the final child weight (Section 5.7). RTI will administer quality assessment and control procedures throughout the process as discussed in Section 1.

5.1 Adjustment for Adult Nonresponse

We classify households with children into two groups based on whether a child-first methodology was used. A child-first household is one where the screener respondent is not the selected adult participant but they are the legal guardian of the adult participant's children. A household with children not employing the child-first methodology (a child-second household) is one where the screener respondent is either the selected adult participant or not a legal guardian of the adult participant's children. Child selection occurred in the screener for the child-first households. By contrast, selection of a child occurred in *Section G* of the adult questionnaire for the child-second households; thus, sampling only occurs for these cases if the adult completes *Section G*. Note that the child-first methodology was only available for households contacted via a landline number because the screener respondent and adult participant were automatically one in the same for cell phone cases.

The input values for the child weights are linked to the probability of selection for their parent. Consequently, RTI will adjust the input values to address adult questionnaire nonresponse (specifically *Section G*) for the child-second households, owing to the dependence of that information on child-level sampling. Specifically, the adult-level input weights will be constructed as follows:

$$AW8_{Thij} = \begin{cases} AW2_{Thij}, & \text{child-first household} \\ AW2_{Thij} \times AA8_{hij}, & \text{child-second household and adult responds} \\ 0, & \text{no children, or child-second household with no adult response} \end{cases} \quad (5.1)$$

a function of the adult base weight, $AW2_{Thij}$ in (4.4), that has been adjusted for cases not subsampled for Phase 2; and a nonresponse adjustment, $AA8_{hij}$, calculated using information available for adult respondents and nonrespondents in households with children. See Table 4.1 for a candidate list of variables. We will use the same techniques discussed for the screener nonresponse adjustment (Section 3.6) for the child weight inputs.

5.2 Base Weights

The child-level base weights, conditional on the selection of the parent, are calculated as the inverse selection probability by age group. Let n_{1ij} represent the number of eligible children ages 0-5 years of the adult respondent and n_{2ij} represent the number of children ages 6-11 years of the adult respondent within household i. We define the probability of selection as follows:

$$CA0_{Thij} = \begin{cases} 2n_{1Thij} \big/ \left(2n_{1Thij} + n_{2Thij}\right), & \text{children ages 0-5 years} \\ n_{2Thij} \big/ \left(2n_{1Thij} + n_{2Thij}\right), & \text{children ages 6-11 years} \end{cases} \quad (5.2)$$

noting that children ages 0-5 years have twice the likelihood of being selected for the study by design. The resulting child-level base weight is expressed as

$$CW0_{Thij} = \begin{cases} AW8_{Thij} / CA0_{Thij}, & \text{household with one parent} \\ \left(AW8_{Thij} / CA0_{Thij}\right) \times 0.5, & \text{household with two parents} \end{cases} \quad (5.3)$$

with terms defined above. For households containing two parents of the selected child, we adjust the weight to account for twice the probability of selection.

5.3 Nonresponse Adjustment

Prior to all nonresponse adjustments, we evaluate data for differential patterns of refusals using correlational statistics and CART. Detectable differences in nonresponse patterns may suggest the need for different models to lower nonresponse bias.

Nonresponse could occur at two points during recruitment for the child interview—parental permission and the interview. We will calculate one or more nonresponse adjustments based on the pattern(s) of nonresponse of the form:

$$CW1_{Thij} = \begin{cases} CW0_{Thij} \times CA1_{Thij}, & \text{child-interview respondents} \\ 0, & \text{child-interview respondents} \end{cases} \quad (5.4)$$

where $CA1_{Thij}$ is the nonresponse adjustment(s) calculated using SUDAAN's PROC WTADJUST. See Table 4.1 for a candidate list of variables known for respondents and nonrespondents. RTI will investigate the use of adult questionnaire responses for inclusion in the model.

5.4 Calibration Adjustment to NHIS

We will calibrate the sample type (LS, CP) child-level weights using the same methodology as implemented for the adult weight adjustment (Section 4.4) for consistency. The calibrated weights will be calculated as:

$$CW2_{Thij} = \begin{cases} CW1_{Thij} \times CA2_{Lhij} & \text{landline-only household} \\ CW1_{Thij} \times CA2_{DLhij} & \text{dual-use landline sample} \\ CW1_{Thij} \times CA2_{DChij} & \text{dual-use cell sample} \\ CW1_{Thij} \times CA2_{Chij} & \text{cell phone-only household} \end{cases} \quad (5.5)$$

where the adjustment factors for landline-only users ($CA2_{Lhij}$), cell phone-only users ($CA2_{Chij}$), and dual users ($CA2_{DLhij}$ and $CA2_{DChij}$) will be calculated with SUDAAN's PROC WTADJUST.

5.5 Composite Factor

As with the adult weight, the calibrated weights for the dual-use households are twice the size of the population values and require adjustment. The resulting composite weight will be constructed as:

$$CW3_{hij} = \begin{cases} CW2_{Thij} & \text{landline-only users} \\ CW2_{Thij} \times \lambda_C & \text{dual users, landline frame} \\ CW2_{Thij} \times (1 - \lambda_C) & \text{dual users, cell frame} \\ CW2_{Thij} & \text{cell phone-only users} \end{cases} \quad (5.6)$$

where $0 \leq \lambda_C \leq 1$ is the composite factor for the child (proxy) dual-use respondents. RTI will investigate various methods to determine λ_C as discussed for λ_A, the adult composite factor. We will allow compositing factors to vary across important analytic domains (including $\lambda_C \neq \lambda_A$) provided that any suggested differences do not affect precision or bias for other key objectives. The sum of the calibrated weights in (5.6) will be verified against the control totals used in the original adjustment in (5.5).

5.6 Calibration Adjustment to Department of Finance Projections

RTI will calibrate the composite weights, $CW3_{ij}$ in (5.6), to the eligible population projections (sans CHIS-ineligible residences) supplied by the State of California's Department of Finance. We will calculate the calibrated weight,

$$CW4_{hij} = CW3_{hij} \times CA4_{hij}, \quad (5.7)$$

as a function of the composite weight and a calibration adjustment factor $CA4_{Thij}$ obtained from SUDAAN PROC WTADJUST. Model covariates will mimic those used for raking in prior rounds of CHIS in addition to variables identified through subsequent analyses. We will verify that the sum of the calibrated weights reproduces the projected estimates before proceeding with the final step.

5.7 Child Analysis Weight

As with the household-level weight (Section 3.8), we will make any final adjustments to the weights based on a multistep quality review. Designating the adjustments as $CA7_{ij}$, the final child analysis weight will be defined as

$$CHW_{hij} = CW4_{hij} \times CA5_{hij} \quad (5.8)$$

RTI will discuss additional weighting refinements with UCLA based on experiences with prior rounds of the study.

6. Teen One-Year Weights

Teens, ages 12 to 17, of the randomly chosen adult are eligible for the study. In contrast to the child (proxy) interview, we recruit one randomly chosen teen to conduct an interview only after receiving permission from a parent.

Below, we describe our proposed plan for calculating a teen analysis weight for analyzing an annual CHIS data file. Steps to calculate the teen weights follow many of those specified for the child weight. Specifically, we define the input values for the teen weights in Section 6.1 that are then adjusted to account for the teen sampling (Section 6.2). We describe in the next sections two (or more) nonresponse adjustments applied to the weights; Section 6.3 addresses lack of parental permission and Section 6.4 nonresponse from the teens themselves. This discussion is followed by a calibration adjustment to population control totals in Section (6.5). The composite factor discussed for the child weight is briefly mentioned in Section 6.6 to combine dual users (landline and cell) selected from both frames. We propose a final calibration adjustment in Section 6.7 with the idea that additional refinements may be necessary to create the final teen weight (Section 6.8). RTI will administer quality assessment and control procedures throughout the process as discussed in Section 1.

6.1 Adjustment for Adult Nonresponse

The input values for the teen weights are linked to the probability of selection for their parent as with the child weight. Consequently, RTI will adjust the input values to address adult questionnaire nonresponse to *Section G*. The adult-level input values proposed for the child weights is used again and is repeated for convenience:

$$AW8_{Thij} = \begin{cases} AW2_{Thij}, & \text{child-first household} \\ AW2_{Thij} \times AA8_{Thij}, & \text{child-second household and adult responds} \\ 0, & \text{no children, or child-second household with no adult response} \end{cases} \quad (6.1)$$

a function of the adult base weight, $AW2_{Tij}$ in (4.4), that has been adjusted for cases not subsampled for Phase 2; and a nonresponse adjustment, $AA8_{ij}$, calculated using information available for adult respondents and nonrespondents in households with children. See Table 4.1 for a candidate list of variables. We will use the same techniques discussed for the screener nonresponse adjustment (Section 3.6) for the child weight inputs.

6.2 Base Weights

The teen base weights, conditional on the selection of the parent, are calculated as the inverse selection probability. Unlike the sampling for children, teens are selected with equal probability. The resulting teen base weight is calculated as

$$TW0_{Thij} = \begin{cases} AW8_{Thij} \times TCNT_{Thij}, & \text{household with one parent} \\ AW8_{Thij} \times TCNT_{Thij} \times 0.5, & \text{household with two parents} \end{cases} \quad (6.2)$$

a function of $AW8_{fij}$ defined in (6.1) and $TCNT_{fij}$ defining the number of eligible teens linked to the participating adult. For households containing two parents of the selected teen, we adjust the weight to account for twice the probability of selection.

6.3 Adjustment for Missing Parent Permission

Nonresponse could occur at two points during recruitment for the teen interview. The first is parental permission. We will calculate one or more nonresponse adjustments based on the pattern(s) of parental refusal for the teen interview:

$$TW1_{Thij} = \begin{cases} TW0_{Thij} \times TA1_{Thij}, & \text{parental permission} \\ 0, & \text{parental refusal} \end{cases} \quad (6.3)$$

where $TA1_{Thij}$ is the nonresponse adjustment(s) calculated using SUDAAN's PROC WTADJUST. See Table 4.1 for a candidate list of variables known for respondents and nonrespondents.

6.4 Adjustment for Teen Nonresponse

The second source of nonresponse is from the teens themselves. We will again calculate one or more nonresponse adjustments based on the pattern(s) of nonresponse of the form:

$$TW2_{Thij} = \begin{cases} TW1_{Thij} \times TA2_{Thij}, & \text{responding teen} \\ 0, & \text{nonresponding teen} \end{cases} \quad (6.4)$$

where $TA2_{Thij}$ is the nonresponse adjustment(s) calculated using SUDAAN's PROC WTADJUST. In addition to the adult candidate variables shown in Table 4.1, RTI will investigate teen variables or adult questionnaire responses to include in the model to assist in reducing nonresponse bias.

6.5 Calibration Adjustment to NHIS

We will calibrate sample type-specific (LS, CP) teen weights using the same methodology as implemented for the adult weight adjustment (Section 4.4). The calibrated weights will be calculated as:

$$TW3_{Thij} = \begin{cases} TW2_{Thij} \times TA3_{Lhij} & \text{landline-only household} \\ \text{TW2}_{Thij} \times TA3_{DLhij} & \text{dual-use landline sample} \\ TW2_{Thij} \times TA3_{DChij} & \text{dual-use cell sample} \\ TW2_{Thij} \times TA3_{Chij} & \text{cell phone-only household} \end{cases} \quad (6.5)$$

where the adjustment factors for landline-only users ($TA3_{Lhij}$), cell phone-only users ($TA3_{Chij}$), and dual users ($TA3_{DLhij}$ and $TA3_{DChij}$) will be calculated with SUDAAN's PROC WTADJUST.

6.6 Composite Factor

The composite weight, calculated to down-weight the dual-use estimates generated with $TW3_{ij}$, will be constructed as:

$$TW4_{hij} = \begin{cases} TW3_{Thij} & \text{landline-only users} \\ TW3_{Thij} \times \lambda_T & \text{dual users, landline frame} \\ TW3_{Thij} \times (1 - \lambda_T) & \text{dual users, cell frame} \\ TW3_{Thij} & \text{cell phone-only users} \end{cases} \quad (6.6)$$

where $0 \le \lambda_T \le 1$ is the composite factor for the teen respondents in the dual-use households. RTI will investigate various methods to determine λ_T as discussed for λ_A, the adult composite factor. We will allow compositing factors to vary across important analytic domains (including $\lambda_T \neq \lambda_C \neq \lambda_A$) provided that any suggested differences do not impact precision or bias for other key objectives. The sum of the calibrated weights in (6.6) will be verified against the control totals used in the original adjustment in (6.5).

6.7 Calibration Adjustment to Department of Finance Projections

RTI will calibrate the composite weights to the eligible population projections supplied by the State of California's Department of Finance as discussed for the adult and child weights. We will calculate the calibrated weight,

$$TW5_{hij} = TW4_{hij} \times TA5_{hij}, \quad (6.7)$$

as a function of the composite weight and a calibration adjustment factor $TA5_{hij}$ obtained from SUDAAN PROC WTADJUST. Model covariates will mimic those used for raking in prior rounds of CHIS in addition to variables identified through subsequent analyses. We will verify that the sum of the calibrated weights reproduces the projected estimates before proceeding with the final step.

6.8 Teen Analysis Weight

We will make any refinements to the calibrated weights based on a multistep quality review. Designating the adjustments as $TA6_{ij}$, we define the final teen analysis weight as

$$TNW_{hij} = TW5_{hij} \times TA6_{hij} \quad (6.8)$$

RTI will discuss additional weighting refinements with UCLA based on experiences with prior rounds of the study.

7. Imputation

RTI will impute plausible values in place of missing responses for variables used to construct the weights. We will conduct the imputation task just after assignment to the response status groups has been finalized (see Section 2). **Table 7.1** provides the candidate list of variable subject to imputation process. Additional variables may be added to this list if needed to finalize the analysis weights and only after consulting with UCLA.

We will impute missing values using one of three methodologies. First, we will use logical imputation (assignment) for variables with low item nonresponse, say 10 percent or lower, that have corroborating information within the associated interview or borrowed from another interview within the same household.[7] For example, we may logically impute missing county of residence as the sampling stratum from which we selected the respondent's landline phone number provided that this assignment is consistent with data from other respondents.

Next, after logically imputing all relevant variables, we next turn to the remaining weighting variables with relatively low levels of nonresponse. This level will be set upon examination of the final cleaned data files and in consultation with UCLA. RTI will impute this second set of missing variables using a weighted sequential hot deck methodology (Cox, 1980; Iannacchione, 1982) with PROC IMPUTE in the SUDAAN statistical software (RTI, 2014). RTI will identify imputation classes using CART methodology as with the nonresponse analyses. Lastly, we will use prediction models to impute for any categorical variables with high item nonresponse and for continuous variables.

As with the construction of the weights, RTI will focus extensively on the quality of the imputed values. We will examine the imputed values for internal consistency (e.g., is the imputed value consistent with other responses within the same interview or same household) and for external consistency (e.g., is the distribution of values different before and after imputation). We will also evaluate the stability of the imputed values by comparing the first set against values obtained from subsequent imputation rounds by, for example, changing the seed value for the randomization process. We will also conduct model diagnostics to ensure the highest utility of the models used for prediction.

RTI will detail the imputation process and results in the Imputation Documentation report (Deliverable 9.3).

[7] Correlational analyses will be conducted for households with multiple interviews to confirm the viability of using responses from one interview as covariates to the imputation model for the other interview. For example, we will determine if adolescent responses can be used to inform missing information for the parent.

Table 7.1 Candidate Variables for Imputation: CHIS 2015-2016

Level	Description
Geographic	• County • Los Angeles SPA • Location information for geocoding (see Deliverable 7.1 for details)
Household	• Household tenure • Presence of cell phone service (landline sample) • Presence of landline service (cell phone sample) • Number of adults in household • Number of children in household • Number of teens in household • Poverty status
Person	• Age • Sex • Education • Race/ethnicity (Latino, white, African American, Asian, American Indian, Alaska Native, Native Hawaiian and Other Pacific Islander, Other) • Asian nationality (including Vietnamese, Korean, and Japanese)

8. Variance Estimation

Variance estimation for CHIS comes in two forms. The first is referred to as Taylor Series linearization or *linearization* for short. The analysis weights described in Sections 4-6 along with the design stratum indicator and survey analysis software (e.g., SUDAAN, Stata, SAS Survey, R) are used to generate (weighted) linearization variance estimates. Design effects (variance given the design divided by the variance under a simple random sample) and coefficients of variation (standard error divided by the estimated average) can be calculated to assess the relative precision of any particular estimate.

The second form of variance estimation is replication. There are several benefits noted for replication variance estimation, including the ability to capture the random nature of the adjustments applied throughout the weighting process. Replicate point estimates (e.g., mean) are generated from replicate weights and used in the following general formula to calculate the associated variance for the point estimate:

$$v\left(\hat{\theta}\right) = a\sum_{r=1}^{R}\left(\hat{\theta}_{(r)} - \hat{\theta}\right)^2 \qquad (8.1)$$

where $\hat{\theta}_{(r)}$ is the estimate generated from the r^{th} replicate; $\hat{\theta}$ is the full-sample estimate of a specific form that depends on the variance estimator chosen (e.g., estimate generated using the linearization weight); and a is a constant depending on the replication method chosen. Replicate weights are formed by first adjusting the base weights for the subsampling and then administering all adjustments applied to the linearization weight to the replicates weights. See Wolter (2003) for a detailed discussion of variance estimation.

RTI proposes to use the same methodology as employed in past rounds of CHIS—a paired-unit grouped jackknife (GJK) replication (see, e.g., Valliant, Brick and Dever, 2008). As of this writing, only the CHIS 2011-2012 methods report for variance estimation was available. In this report, 80 GJK replicates were generated; RTI proposes to use the same number of replicates from CHIS 2014 to maintain the same level of precision. We will investigate the need to generate more replicates to address, for example, subgroup analyses.

References

Blumberg, S.J. and Luke, J.V. (2015). Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July–December 2014. Division of Health Interview Statistics, National Center for Health Statistics. URL: http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201506.pdf.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.

Brick, J.M., Flores Cervantes, I., Lee, S., and Norman, G. (2011). Nonsampling errors in dual frame telephone surveys. *Survey Methodology*, **37**(1):1-12.

California Health Interview Survey (CHIS). *CHIS 2011-2012 Methodology Series: Report 5 – Weighting and Variance Estimation*. Los Angeles, CA: UCLA Center for Health Policy Research, 2014.

Chen, P. et al. (2012). National Survey on Drug Use and Health: Person-level sampling, Weight Calibration. Prepared for the Substance Abuse and Mental Health Services Administration as part of the 2012 Methodological Resource Book, February 2014. http://www.samhsa.gov/data/sites/default/files/NSDUH2012MRB-Ammended/NSDUHmrbSamplingWgt2012.pdf

Cox, B. (1980). The weighted sequential hot deck imputation procedure. In Proceedings of the American Statistical Association, Section on Survey Research Methods, 721-726. Available at www.amstat.org/sections/srms/proceedings/papers/1980_152.pdf.

Iannacchione, V.G. (2003). Sequential weight adjustments for location and cooperation propensity for the 1995 National Survey of Family Growth. *Journal of Official Statistics*, **19**:31-43.

Folsom, R.E., & A.C. Singh (2000). "The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and post-stratification." *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 598-603. Available at http://www.amstat.org/sections/SRMS/Proceedings/papers/2000_099.pdf.

Kim, J., Li, J., and Valliant, R. (2007). Cell collapsing in poststratification. *Survey Methodology*, **33**:139-150.

Kott, P.S. (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. *Survey Methodology*, **32**(2): 133-142.

Kott P.S. & Liao D (2012). Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine. *Survey Research Methods*, **6**(2): 105-111.

Kott, P. (2014). Calibration Weighting When Model and Calibration Variables Can Differ. In Contributions to Survey Statistics - ITACOSM 2013 Selected Papers (pp. 1-18). Cham: Springer, Contributions to Statistics.

Research Triangle Institute (2012). *SUDAAN Language Manual, Volumes 1 and 2, Release 11*. Research Triangle Park, NC: Research Triangle Institute.

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.

RTI International. (2012). SUDAAN: Statistical software for weighting, imputing, and analyzing data, Release 11. URL: http://www.rti.org/sudaan/.

Särndal C., Swensson B., & Wretman J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Valliant, R., Brick, J.M., and Dever, J. (2008). Weight adjustments for the grouped jackknife variance estimator. *Journal of Official Statistics*, **24**, 469-488.

Valliant, R., Dever, J., and Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*, New York: Springer.Wolter, K.M. (2003). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Endnotes

[1] Trimming should only occur within frame or within sample type if it is needed. RTI will consult with UCLA on our evaluation of the need for weight trimming. We will document the use of any weight trimming, including where it occurred in the process and the number of cases affected.

Please know that weight trimming should be used if absolutely needed because this procedure can introduce bias into the estimates.

[2] UCLA currently recommends "health insurance" as one of the variables used for trend and other analyses. They will identify other variables for consideration.

[3] RTI intends to deliver to UCLA all weight components and subsequent interim weights so that a complete evaluation may be conducted. We use this level of information for our own QC as well.

RTI proposes to include only the final weights on the interview data files delivered under Task 8. We will submit the complete list of weighting variables (interim and final) on a data file to accompany the Weighting Documentation report (Deliverable 9.2) unless otherwise specified. RTI will work with UCLA to have the data structured in a cost-effectively way to facilitate review.

[4] RTI obtains all landline sample cases from the vendors regardless of their purge status. We screen the numbers again using our internal systems. First, we screen to determine if any of the numbers are now registered to a cell phone (i.e., ported). Next, we send the non-ported numbers through an auto dialer to determine working status. Only those we have found to be working are released. We continue to evaluate where the vendor status can be used outright without a second screening process but to date we only field our designated working numbers.

[5] Our modeling approach via SUDAAN's PROC WTADJUST [see endnote 6] is a generalization of the weighting class and poststratification used in the past (see, e.g., sections 14.2 and 14.3 in Valliant et al., 2013; Särndal et al., 1992). In other words, our recommended approach is related to techniques used previously and affords us the opportunity to include additional variables in the adjustment model.

Weighting class adjustment procedures (WCA) use categorical (or categorized) variables to form weighting classes. These classes must be evaluated to ensure they contain sufficient numbers of cases. Sufficiency is somewhat subjective but many researchers use 30 as the minimum number respondent cases. This is especially problematic for surveys with low response. If a class is deemed too small, then researchers must determine which of the "surrounding" classes should be collapsed to alleviate the problem. This determination takes time (and can be somewhat subjective) because collapsing classes inappropriately can introduce bias into the estimates. Hence, WCA typically includes only a small number of covariates.

Ideally, the covariates used in a nonresponse WCA should be correlated with patterns of nonresponse and with key analysis variables. The former is the most common type of variable used (identified, for example, via a classification and regression tree [CART] analysis) since this information must be known for respondents and nonrespondents.

Using a general model-based approach allows for the inclusion of additional covariates even continuous variables and complex interactions not feasible within a WCA setting. Expanding the set of covariates can reduce nonresponse bias but this must be done with caution as to not introduce excessive variation in the adjustment factors.

From our read of the methods report, there was a large amount of time spent collapsing individual weighting classes. Through a model-based approach, we can set parameters to handle this work directly. We can compare the results of our analyses against the details provided in the 2013-2014 report.

RTI will estimate the impact to the budget and schedule for creating "CHIS 2013-2014 like" weights with the CHIS 2015 data using the latest methods report, and will report our findings to UCLA. However, note that the design was changed in CHIS 2015 so any differences in the weighting approach is confounded with a series of changes to the study.

Note that including many dimensions in the model may reduce bias but at the expense of the standard errors if the variation in the weights is excessive and not correlated with the variable of interest. Having fewer dimensions may reduce the weight variation at the expense of bias. A nonresponse bias analysis may inform which of several methods appear to be more effective. We will conduct such an analysis throughout our weighting process.

[6] WTADUST, a procedure within SUDAAN® (Research Triangle Institute 2012), uses a generalized exponential model with user-defined (if desired) upper and lower bounds on the weight adjustments to control for excessive variation. Note that this is a form of weight smoothing.

Covariates in this model are those identified for nonresponse adjustment or final adjustment to population controls. The generalized nature of the underlying model allows for specific adjustment procedures such as:

- weighting class adjustment or poststratification adjustment (all interaction terms from a set of categorical variables are included in the model, excluding the model intercept)
- iterative proportional fitting or raking adjustment (marginal covariates)
- calibration (combination of covariates used for poststratification and raking).

Output from the model is a set of predicted probabilities that are applied inversely to the "input" weights to adjust for nonresponse or to align with population values. Details on the methodology are found in chapter 24 of the SUDAAN Language Manual (Research Triangle Institute 2012), Folsom & Singh (2000), Kott & Liao (2012) and Kott (2014).

The CHIS stats team has extensive experience in using this procedure and with the underlying methodology.

[7] CART stands for classification and regression tree analysis. This is one of the techniques used to assess associations within the data. We also intend to examine pairwise correlations to develop parsimonious adjustment models. We will compare our results with past rounds of CHIS to identify and discuss with UCLA recommended changes.

[8] Estimates for the entire nation are not appropriate for California-specific weight adjustments. Based on information only available to date in the 2011-12 report, we assume that wireless population estimates at the time were preliminary and only available at the regional level (e.g., western U.S.). Subsequent estimates were released at the state-level, showing variation among the states within the West region. Additionally, there has been at least one release of estimates for large counties in California, along with sub-state regions.

We will focus on the California-specific estimates for our adjustment procedures and will investigate the availability (and viability) of using model-based estimates for smaller areas within CA. For example, current documentation for CHIS indicates that control total information was obtained from the following sources. RTI will investigate there use for CHIS 2015-2016.

- California Department of Finance (CA-DOF) Population Predictions and Estimates
- Census 2010 Files (estimates used to remove group quarters from CA-DOF predictions)
- American Community Survey for California Public-use Microdata (estimates by tenure and education status by region)
- The National Health Interview Survey (wireless and dual-use estimates).

Chapter 3: Additional Information on Weighting and Imputation

Table 3-1 below describes the response status groups used for weighting in CHIS 2015. This table is modelled on Table 4-1 in CHIS 2013-2014.

Table 3-1. Response Status

Response Status Group	Description
Eligible Respondent	person sampled through eligible phone number and completes interview
Eligible Nonrespondent	person sampled through eligible phone number and either refuses to participate or does not answer a sufficient number of items in the questionnaire (i.e., not a usable for analyses)
Ineligible sample member	business, group quarters, non-working, household without at least one adult (18 years of age), cell phone link only to person(s) less than 18 years of age, other non-residential
Unknown eligibility	all other cases

Note: Information in table is from Section 2.1 of DLV 9.2; not shown as an explicit table. This table corresponds to CHIS 2013-2014 Report 5, Table 4-1.

Dimensions used in adult weighting adjustments

Table 3-2 below shows the dimensions used in development of the adult weights.

Table 3-2. Variables Used in Nonresponse Adjustment Applied to the Adult Weights: CHIS 2015

Level	Description
Geographic/ Sampling	· Frame type · Design stratum / county · Hispanic surname · Asian surname · Address with sampled number
Household	· Number of kids 0-17 in home · Number of adults · Child-first household
Person	· Is the person on a cell phone

* Information must be available for respondents and nonrespondents to be used in the adjustment model.

Note: This table modelled after CHIS 2013-2014 Table 4-2 "Variables used for the creation of nonresponse adjustment cells for the adult weights"

Variables and Calibration Levels for CHIS 2015 Adult Weights

Table 3-3 below shows the variables (and levels for each variable) used in the calibration adjustment for adult weights.

Table 3-3. Variables and Levels for Calibration Adjustment Applied to the Adult Weights: CHIS 2015-2016

Group	Characteristic	Levels	Description
Geography	Reported Stratum (44)[1]	1	Los Angeles
		2	San Diego
		3	Orange
		4	Santa Clara
		5	San Bernardino
		6	Riverside
		7	Alameda
		8	Sacramento
		9	Contra Costa
		10	Fresno
		11	San Francisco
		12	Ventura
		13	San Mateo
		14	Kern
		15	San Joaquin
		16	Sonoma
		17	Stanislaus
		18	Santa Barbara
		19	Solano
		20	Tulare
		21	Santa Cruz
		22	Marin
		23	San Luis Obispo
		24	Placer
		25	Merced
		26	Butte
		27	Shasta
		28	Yolo
		29	El Dorado
		30	Imperial
		31	Napa
		32	Kings
		33	Madera

		34	Monterey
		35	Humboldt
		36	Nevada
		37	Mendocino
		38	Sutter
		39	Yuba
		40	Lake
		41	San Benito
		42	Tehama-Glenn-Colusa
		43	Del Norte-Siskiyou-Lassen-etc.
		44	Tuolumne-Calaveras-Amador-Inyo-etc.
Geography	Reported Los Angeles SPA [1]	1	Antelope Valley
		2	San Fernando Valley
		3	San Gabriel Valley
		4	Metro
		5	West
		6	South
		7	East
		8	South Bay
	Reported San Diego HSR [1]	1	North Coastal
		2	North Central
		3	Central
		4	South
		5	East
		6	North Inland
	Sub-state Region [2]	1	Northern & Sierra Counties (Butte, Shasta, Humboldt, Lake, Mendocino, Yuba, Nevada, Sutter, Colusa, Glenn, Tehama, Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity, Alpine, Amador, Calaveras, Inyo, Mariposa, Mono. Tuolumne)
		2	Greater Bay Area (Santa Clara, Alameda, Contra Costa, San Francisco, San Mateo, Sonoma, Solano, Marin, Napa)
		3	Sacramento Area

			(Sacramento, Placer, Yolo, El Dorado)
		4	San Joaquin Valley (Fresno, Kern, San Joaquin, Stanislaus, Tulare, Merced, Kings, Madera)
		5	Central Coast (Ventura, Santa Barbara, Santa Cruz, San Luis Obispo, Monterey, San Benito)
		6	Los Angeles
		7	Other Southern California (San Diego, Orange, San Bernardino, Riverside, Imperial)
	State excluding LA, SD [2]	1	Remainder of CA
Person	Age (categorized)	1	18-24
		2	25-29
		3	30-39
		4	40-49
		5	50-64
		6	65 and over
	Sex	1	Male
		2	Female
	Adult education	1	Less than High School
		2	High School grad or GED
		3	At least some college
	Race/ethnicity	1	Latino 18 years old or older
		2	Non-Latino White 18 years old or older
		3	Non-Latino African American 18 years+
		4	Non-Latino American Indian 18 years+
		5	Non-Latino Asian 18 years old or older
		6	Non-Latino Native Hawaiian 18 years+
		7	Non-Latino 2+ races 18 years+
	Asian nationality	1	Non-Latino Chinese
		2	Non-Latino Korean
		3	Non-Latino Filipino

		4	Non-Latino Vietnamese
		5	Non-Latino Japanese
		6	Non-Latino Other Asian
	Nontelephone	1	HH owned, 18 to 30, <= HS, all adults
		2	HH owned, 31 to 64, <= HS, all adults
		3	HH owned, 65 and older, <= HS, all adults
		4	HH owned, 18 to 30, > HS, all adults
		5	HH owned, 31 to 64, > HS, all adults
		6	HH owned, 65 and older, > HS, all adults
		7	HH rented, 18 to 34, <= HS, all adults
		8	HH rented, 35+ yrs, <= HS, 1 adult
		9	HH rented, 35+ yrs, <= HS, 2 adults
		10	HH rented, 18 to 34, > HS, all adults
		11	HH rented, 35+ yrs, > HS, 1 adult
		12	HH rented, 35+ yrs, > HS, 2 adults
	NHIS Calibration	1	Cell phone only HH
		2	Dual phone user HH
		3	Landline only HH

[1] Reported stratum (county) information.

[2] Variable derived from Reported stratum (county) information.

Note: Corresponds to Table 7-1 in CHIS 2013-2014.

Item Nonresponse and Imputation Rates for Self-reported Sex and Age

As reported in the weighting/imputation memo above, variables that were used by RTI in the weighting process were first imputed by RTI. Table 3-4 shows the item nonresponse rate for self-reported age.

Table 3-4. Item Nonresponse for Self-reported and Age by Person Type and Interview Mode

			Interview Mode			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct[1]	n	pct[1]	n	pct[1]
SRSEX (Self-reported sex)						
Adult	5	0.0	5	0.1	0	0.0
Child	5	0.2	1	0.1	4	0.4
Teen	1	0.1	0	0.0	1	0.3
SRAGE (Self-reported age)						
Adult	150	0.7	41	0.4	109	1.0
Child	13	0.6	3	0.2	10	1.0
Teen	8	1.1	3	0.8	5	1.3

[1] Unweighted percent of imputed records among respondents in Table 5-1 by mode and person type.

Note: Corresponds to Table 8-2 in CHIS 2013-2014.

Item Nonresponse and Imputed Rate for Self-reported Education

Table 3-5 shows the item nonresponse rate for self-reported education.

Table 3-5. Item Nonresponse for Self-reported Educational Attainment of the Adult by Person Type and Interview Mode

			Interview Mode			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct	n	pct	n	pct
SREDUC (Self-reported educational attainment)						
Reported	21,028	100.0	9,769	46.5	11,259	53.5
Imputed – Hot Deck	6	0.0	4	0.0	2	0.0
Total	21,034	100.0	9,773	46.5	11,261	53.5

[1] Unweighted percent of cases within interview mode and variable.

Note: Corresponds to Table 8-4 in CHIS 2013-2014.

Pre- and Post-imputation Distribution Comparisons

Table 3-6 below shows the unweighted distributions of household-level variables pre- and post-imputation to assess the effect of the imputation conducted on those variables.

Table 3-6. Unweighted Distribution of Household-level Variables Pre- and Post-Imputation

ELIG_KID_0_5 (Self-reported number of eligible children age 0-5)

	Pre-imputation		Post-imputation		Differ
Number	n	%	n	%	%
0	19,500	92.72	19,503	92.72	0.0
1	1,052	5.00	1,053	5.01	0.0
2	421	2.00	421	2.00	0.0
3	54	0.26	54	0.26	0.0
4	3	0.01	3	0.01	0.0

ELIG_KID_6_11 (Self-reported number of eligible children age 6-11)

	Pre-imputation		Post-imputation		Differ
Number	n	%	n	%	%
0	19,107	90.86	19,110	90.85	-0.0
1	1,326	6.31	1,327	6.31	0.0
2	505	2.40	505	2.40	0.0
3	85	0.40	85	0.40	0.0
4	5	0.02	5	0.02	0.0
5	2	0.01	2	0.01	0.0

ELIG_TEEN (Self-reported number of eligible teens)

	Pre-imputation		Post-imputation		Differ
Number	n	%	n	%	%
0	18,925	89.99	18,925	89.97	-0.0
1	1,516	7.21	1,519	7.22	0.0
2	514	2.44	514	2.44	0.0
3	64	0.30	64	0.30	0.0
4	10	0.05	10	0.05	0.0
5	1	0.00	1	0.00	0.0
6	1	0.00	1	0.00	0.0

PARENT_CHILD_HH (Self-reported number of parents related to the selected child)

Number	Pre-imputation n	Pre-imputation %	Post-imputation n	Post-imputation %	Differ %
0	18,256	86.81	18,256	86.79	-0.0
1	2,397	11.40	2,400	11.41	0.0
2	377	1.79	378	1.80	0.0

PARENT_TEEN_HH (Self-reported number of parents related to the selected teen)

Number	Pre-imputation n	Pre-imputation %	Post-imputation n	Post-imputation %	Differ %
0	18,925	89.99	18,925	89.97	-0.0
1	1,787	8.50	1,789	8.51	0.0
2	319	1.52	320	1.52	0.0

HASCELL (Presence of a cell phone)

Presence	Pre-imputation n	Pre-imputation %	Post-imputation n	Post-imputation %	Differ %
Yes	18,492	88.39	18,576	88.31	-0.0
No	2,428	11.61	2,458	11.69	0.0

HASLANDLINE (Presence of a landline phone)

Presence	Pre-imputation n	Pre-imputation %	Post-imputation n	Post-imputation %	Differ %
Yes	15,085	72.76	15,209	72.31	-0.4
No	5,648	27.24	5,825	27.69	0.4

CALLINTENSITY (Self-reported phone intensity)

Presence	Pre-imputation n	Pre-imputation %	Post-imputation n	Post-imputation %	Differ %
Cell Only	5,825	32.82	5,825	27.69	-5.1
Dual User - Mostly Cell	2,792	15.73	3,952	18.79	3.0
Dual User - Equal Use	3,593	20.24	4,791	22.78	2.5
Dual User - Mostly Landline	3,083	17.37	4,008	19.05	1.6
Landline Only	2,458	13.85	2,458	11.69	-2.1

SRTENR (Self-reported tenure)

Presence	Pre-imputation n	Pre-imputation %	Post-imputation n	Post-imputation %	Differ %
Own	12,153	58.49	12,284	58.4	-0.0
Rent	7,833	37.7	7,946	37.78	0.0
Other Arrangement	793	3.82	804	3.82	0.0

Poverty (Self-reported poverty status)

Presence[1]	Pre-imputation n	Pre-imputation %	Post-imputation n	Post-imputation %	Differ %
1 - <= 50% FPL	847	4.03	847	4.03	0.0
2 - 50% < FPL <= 100%	1,778	8.46	1,778	8.45	0.0
3 - 100% < FPL <= 133%	1,382	6.57	1,382	6.57	0.0
4 - 133% < FPL <= 200%	1,834	8.72	1,834	8.72	0.0
5 - 200% < FPL <= 300%	2,312	10.99	2,313	11.00	0.0
6 - 300% < FPL <= 400%	1,760	8.37	1,763	8.38	0.0
7 - FPL > 400%	7,288	34.66	7,290	34.66	0.0
8 - Unknown	3,827	18.20	3,827	18.19	-0.0

[1] FPL = Federal Poverty Level

Note: Corresponds to Table 8-4 in CHIS 2013-2014.

Item Nonresponse for Self-reported Race and Ethnicity by Interview Type

Table 3-7 shows the item nonresponse rate for self-reported race and ethnicity by interview type (i.e., adult, child, teen).

Table 3-7. Item Nonresponse for Any Self-reported Race Value and Ethnicity by Interview Type

			Interview Mode			
	All Modes		Cell		Landline	
Variable and Source of Data	n	pct	n	pct	n	pct
One or more imputed Race values						
Adult	605	2.9	366	3.7	239	2.1
Child	123	5.7	66	5.8	57	5.6
Teen	59	7.8	35	9.4	24	6.3
SRH (Self-reported Latin ethnicity)						
Adult	121	0.6	50	0.5	71	0.6
Child	123	5.7	66	5.8	57	5.6
Teen	59	7.8	35	9.4	24	6.3

[1] Unweighted percent of imputed records among respondents in Table 5-1 by mode and person type.

Note: Corresponds to Table 8-6 in CHIS 2013-2014.

Item Nonresponse for Phone Type Presence and Use

Table 3-8 shows the item nonresponse and imputation rate to questions that ask about the type of phones that can reach the respondent and their use. These variables are important for weighting dual-frame (landline and cell) RDD samples.

Table 3-8. Item Nonresponse for Presence of Cell Phone, Presence of Landline Phone, and Type of Phone Usage by Interview Mode

Variable and Source of Data	All Modes		Interview Mode: Cell		Interview Mode: Landline	
	n	pct[1]	n	pct[1]	n	pct[1]
HASCELL (Presence of a cell phone)						
Reported	20,920	99.5	9,773	100.0	11,147	99.0
Imputed – Hot Deck	114	0.5	*na*	*na*	114	1.0
Total	21,034	100.0	9,773	100.0	11,261	100.0
HASLANDLINE (Presence of a landline phone)						
Reported	20,733	98.6	9,472	96.9	11,261	100.0
Imputed – Hot Deck	301	1.4	301	3.1	*na*	*na*
Total	21,034	100.0	9,773	100.0	11,261	100.0
CALLINTENSITY (Self-reported phone intensity)						
Reported	17,747	84.4	6,651	68.1	11,096	98.5
Imputed – Hot Deck	3,287	15.6	3,122	32.0	165	1.5
Total	21,034	100.0	9,773	100.0	11,261	100.0

Note: *na* = not applicable.

[1] Unweighted percent of cases within interview mode and variable.

Note: Corresponds to Table 8-12 in CHIS 2013-2014.

Listing of Variables used in Geocoding

Table 3-9 is a simple list of variables used in geocoding and their attributes.

Table 3-9. Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
1	BASEID	Char	8	$8.00	$8.00	Unique Record ID
7	BEST_ZIP	Num	8			BEST_ZIP
12	GEO_BLKGRP	Num	8			GEO_BLKGRP
13	GEO_FIPS	Num	8			GEO_FIPS
9	GEO_FIPS_CNTY	Num	8			GEO_FIPS_CNTY
11	GEO_TRACT	Num	8			GEO_TRACT
14	I_BEST_ZIP	Num	8			I_BEST_ZIP
6	LATITUDE	Num	8			LATITUDE
5	LONGITUDE	Num	8			LONGITUDE
3	MATCHSTATUS	Char	14	$14.00	$14.00	MATCHSTATUS
4	SCORE	Num	8			SCORE
8	STATE_FIPS	Num	8			STATE_FIPS
10	STCOFIPS	Num	8			STCOFIPS
2	voxco_CASEID	Char	10	$10.00	$10.00	resRespondent
15	zeropop	Num	8			Flag for zero population

Note: Corresponds to Table 8-13 in CHIS 2013-2014. For CHIS 2015, this is not an explicit table but is described in the last section of DLV 7.4 (Geocoding Plan).

Weighting Adjustment Magnitudes

Tables B-1 through B-7 in past cycle reports were collapsed into 4 tables (3-10 through 3-13). These tables describe the magnitude of the weighting adjustment applied to the various samples.

Table 3-10 below displays the adjustment magnitudes for landline, list, and cell frames at various stages of adjustment. This table corresponds to Table B-1 in CHIS 2013-2014 methodology reports.

Table 3-10. Screener interview (households) weighting adjustments by sample type

	Sampling Frame		
Survey Weight Statistics[1]	**Landline**	**List**	**Cell**
1. Base weight			
1.1 Sample size	271,032	8,274	206,429
1.2 Sum of weights	65,436,912.99	324,273.27	211,735,600.00
1.3 Coefficient of variation	61.74	62.54	75.89
2. Surname list sampling adjustment			
2.1 Sample size	271,032	8,274	206,429
2.2 Sum of weights	65,436,912.99	162,136.63	211,735,600.00
2.3 Coefficient of variation	61.74	62.54	75.89
3. Adjustment for multiple periods, frames and vendors			
3.1 Sample size	271,032	8,274	206,429
3.2 Sum of weights	17,738,575.64	43,998.67	52,923,300.00
3.3 Coefficient of variation	61.20	62.31	74.58
4. Phase 2 subsampling adjustment			
4.1 Sample size	225,200	5,117	158,262
4.2 Sum of weights	18,049,934.87	45,494.55	53,866,475.61
4.3 Coefficient of variation	84.27	98.22	108.94
4.4 Mean non-zero adjustment factor	1.23	1.67	1.32
5. Unknown residential status adjustment			
5.1 Sample size			
a. Known residential	25,063	1,040	34,328
b. Unknown residential	200,137	4,077	123,934
5.2 Sum of weights	2,586,783.51	9,541.86	19,545,237.88
5.3 Coefficient of variation	85.74	97.83	112.87
5.4 Mean non-zero adjustment factor	1.13	1.13	1.47

6. Unknown eligibility adjustment			
6.1 Sample size			
a. Residential, Eligible	24,389	1,026	21,521
b. Residential, Unknown eligibility	674	14	12,807
6.2 Sum of weights	2,586,593.06	9,731.79	19,545,237.88
6.3 Coefficient of variation	85.10	94.70	91.92
6.4 Mean non-zero adjustment factor	1.05	1.05	1.96
7. Screener nonresponse adjustment			
7.1 Sample size			
a. Screener respondents	22,137	937	16,620
b. Screener nonrespondents	2,252	89	4,901
7.2 Sum of weights (screener respondents)	2,586,590.95	9,731.79	19,545,237.88
7.3 Coefficient of variation	82.30	91.70	87.99
7.4 Mean non-zero adjustment factor	1.16	1.15	1.29

[1] *Statistics after application of each weight adjustment*

Table 3-11 below displays the adjustment magnitudes at various stages for the adult sample weights. This table corresponds to Table B-2 in CHIS 2013-2014 methodology reports

Table 3-11. Extended interview weighting procedures for adult interviews by sample type

	Sampling Frame		
Survey Weight Statistics[1]	**Landline**	**List**	**Cell**
1. Adult Base Weight			
1.1 Sample size	22,138	937	16,620
1.2 Sum of weights	4,961,093.02	20,197.92	19,696,483.60
1.3 Coefficient of variation	98.81	112.55	89.56
2. Phase 2 Adjustment			
2.1 Sample size	22,138	937	16,620
2.2 Sum of weights	4,910,492.44	19,886.85	19,264,756.30
2.3 Coefficient of variation	99.02	114.05	89.71
3. Nonresponse Adjustment			
3.1 Sample size			
a. Adult respondents	11,211	463	9,360
b. Adult nonrespondents	10,927	474	7,260
3.2 Sum of weights	4,907,491.70	22,887.60	19,264,756.30
3.3 Coefficient of variation	120.67	117.29	89.35
3.4 Mean non-zero adjustment factor	2.19	2.62	2.02
4. Calibration to telephone service			
4.1 Sample size	11,211	463	9,360
4.2 Sum of weights	6,330,625.03	30,351.89	21,837,004.00
4.3 Coefficient of variation	126.81	113.72	90.01
4.4 Mean non-zero adjustment factor	0.78	0.79	0.72
5. Composite weight			
5.1 Sample size	11,211	463	9,360
5.2 Sum of weights	5,053,029.14	24,739.59	15,135,887.50
5.3 Coefficient of variation	138.91	118.83	103.13
5.4 Mean non-zero adjustment factor	1.00	1.00	1.00
6. Trimming Adjustment*			
6.1 Number of Trimmed Records	0	0	4
6.2 Sum of weights	5,053,029.14	24,739.59	15,129,555.00
6.3 Coefficient of variation	138.91	118.83	103.16
6.4 Mean non-zero adjustment factor	1.95	2.99	1.67
7. Calibration adjustment			
7.1 Sum of weights	9,270,736.56	67,053.18	19,744,862.30
7.2 Coefficient of variation	154.94	133.91	111.15
7.3 Mean non-zero adjustment factor	1.95	2.99	1.67

7.4 Mean Weight	826.93	144.82	2,109.49

[1] *Statistics after application of each weight adjustment*

Table 3-12 below displays the adjustment magnitudes for the child weights at various stages of adjustment. This table corresponds to Table B-3 in CHIS 2013-2014 methodology reports.

Table 3-12. Extended interview weighting procedures for child interviews by sample type

	Sampling Frame		
Survey Weight Statistics[1]	**Landline**	**List**	**Cell**
1. Adult nonresponse adjustment			
1.1 Sample size			
a. Adult respondents	1,011	47	1,720
b. Adult nonrespondents	1,557	38	1,455
1.2 Sum of weights	868,433.57	2,574.25	5,678,588.11
1.3 Coefficient of variation	99.82	80.75	92.63
1.4 Mean non-zero adjustment factor	1.64	1.52	2.08
2. Base weights			
2.1 Sample size	2,568	85	3,175
2.2 Sum of weights	945,910.39	2,233.85	6,337,679.74
2.3 Coefficient of variation	105.80	68.69	107.71
3. Child Interview Nonresponse Adjustment			
3.1 Sample size			
a. Child interview respondents	997	36	1,124
b. Child interview nonrespondents	1,571	49	2,051
3.2 Sum of weights	945,910.39	2,233.85	6,337,679.74
3.3 Coefficient of variation	111.82	82.16	104.05
3.4 Mean non-zero adjustment factor	2.19	2.14	2.33
4. Calibration to telephone service			
4.1 Sample size	997	36	1,124
4.2 Sum of weights	942,271.59	2,115.33	4,858,229.56
4.3 Coefficient of variation	123.45	83.52	104.69
4.4 Mean non-zero adjustment factor	1.00	0.95	0.77
5. Composite weight			
5.1 Sample size	997	36	1,124
5.2 Sum of weights	736,802.09	1,603.34	3,524,259.75
5.3 Coefficient of variation	139.38	91.26	119.27
5.4 Mean non-zero adjustment factor	0.75	0.73	0.77
6. Trimming Adjustment			
6.1 Number of Trimmed Records	0	0	6

6.2 Sum of weights	736,802.09	1,603.34	3,507,063.14
6.3 Coefficient of variation	139.38	91.26	119.29
6.4 Mean non-zero adjustment factor	1.00	1.00	1.00
7. Calibration adjustment			
7.1 Sum of weights	1,584,052.89	5,764.65	4,465,051.11
7.2 Coefficient of variation	122.71	114.91	105.39
7.3 Mean non-zero adjustment factor	2.80	4.13	1.72
7.4 Mean Weight	1,588.82	160.13	3,972.47

[1] *Statistics after application of each weight adjustment*

Table 3-13 below displays the adjustment magnitudes for the teen weights at various stages of adjustment. This table corresponds to Table B-4 in CHIS 2013-2014 methodology reports

Table 3-13. Extended interview weighting adjustments for teen interviews by sample type

	Sampling Frame		
Survey Weight Statistics[1]	**Landline**	**List**	**Cell**
1. Base weights			
1.1 Sample size	2,104	105	1,964
1.2 Sum of weights	646,279.70	2,664.31	2,664,287.17
1.3 Coefficient of variation	111.34	121.65	103.83
2. Teen nonresponse adjustment			
2.1 Sample size			
a. Teen respondents	367	20	367
b. Teen nonrespondents	1,737	85	1,597
2.2 Sum of weights	646,279.70	2,664.31	2,664,287.17
2.3 Coefficient of variation	112.83	64.36	130.45
2.4 Mean non-zero adjustment factor	6.27	5.59	6.04
3. Calibration to telephone service			
3.1 Sample size	367	20	367
3.2 Sum of weights	447,509.60	1,854.07	2,391,396.00
3.3 Coefficient of variation	111.18	68.08	114.73
3.4 Mean non-zero adjustment factor	0.70	0.68	0.93
4. Composite weight			
4.1 Sample size	367	20	367
4.2 Sum of weights	334,854.79	1,358.84	1,769,994.78
4.3 Coefficient of variation	112.62	64.45	141.26
4.4 Mean non-zero adjustment factor	0.75	0.76	0.70
5. Trimming Adjustment			
5.1 Number of Trimmed Records	2	0	8
5.2 Sum of weights	330,545.56	1,358.84	1,702,419.61
5.3 Coefficient of variation	110.39	64.45	141.69
5.4 Mean non-zero adjustment factor	1.00	1.00	0.99
6. Calibration adjustment			
6.1 Sum of weights	831,206.03	6,663.84	2,116,173.79
6.2 Coefficient of variation	140.75	128.76	126.27
6.3 Mean non-zero adjustment factor	2.69	5.32	1.96
6.4 Mean Weight	2,264.87	333.19	5,766.14

[1] *Statistics after application of each weight adjustment*