August, 2009

CHIS 2007 Methodology Report Series

Report 5

Weighting and Variance Estimation

CHIS 2007 METHODOLOGY SERIES

REPORT 5

WEIGHTING AND VARIANCE ESTIMATION

August 2009

This report was prepared for the California Health Interview Survey by Ismael Flores Cervantes, Greg Norman, and J. Michael Brick of Westat.

www.chis.ucla.edu

This report describes the weighting and variance estimation methods used in CHIS 2007. This report presents the steps used to create the analytical weights for analyzing the data from the adult, child, and adolescent interviews.

Suggested citation:

California Health Interview Survey. *CHIS 2007 Methodology Series: Report 5 – Weighting and Variance Estimation*. Los Angeles, CA: UCLA Center for Health Policy Research, 2009.

The California Health Interview Survey is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Public Health, the Department of Health Care Services, and the Public Health Institute. Funding for CHIS 2007 came from multiple sources: the California Department of Public Health, the Department of Health Care Services, the California Endowment, the National Cancer Institute, the First 5 California, the Robert Wood Johnson Foundation, the California Department of Mental Health, the California Office of the Patient Advocate, Kaiser Permanente, Blue Shield of California Foundation, LA Care Health Plan, the San Diego County Human and Health Services Agency, and the California Attorney General's Crime and Violence Prevention Center.

PREFACE

Weighting and Variance Estimation is the fifth in a series of methodological reports describing the 2007 California Health Interview Survey. The other reports are listed below.

CHIS is a collaborative project of the University of California, Los Angeles (UCLA) Center for Health Policy Research, the California Department of Public Health, the Department of Health Care Services, and the Public Health Institute. Westat was responsible for data collection and the preparation of five methodological reports from the 2007 survey. The survey examines public health and health care access issues in California. The telephone survey is the largest state health survey ever undertaken in the United States. The plan is to monitor these issues and examine changes over time by conducting surveys in the future.

Methodological Reports

The first five methodological reports for CHIS 2007 are as follows:

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

The reports are interrelated and contain many references to each other. For ease of presentation, the references are simply labeled by the report numbers given above.

This report describes the weighting and variance estimation methods from CHIS 2007. The purpose of weighting the survey data is to permit analysts to produce estimates of the health characteristics for the entire California population and subgroups including counties, and in some cases, cities. This report presents the steps used to create the analytical weights for analyzing the data from the adult, child, and adolescent interviews.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

List of Appendices

TABLE OF CONTENTS (continued)

List of Appendices (Continued)

List of Tables

TABLE OF CONTENTS (continued)

List of Tables (continued)

List of Figures

1. CHIS 2007 DESIGN AND METHODOLOGY SUMMARY

1.1 Overview

The California Health Interview Survey (CHIS) is a population-based telephone survey of California's population conducted every other year since 2001. CHIS is the largest health survey conducted in any state and one of the largest health surveys in the nation. CHIS is based at the UCLA Center for Health Policy Research (CHPR) and is conducted in collaboration with the California Department of Public Health, the Department of Health Care Services, and the Public Health Institute. CHIS collects extensive information for all age groups on health status, health conditions, health-related behaviors, health insurance coverage, access to health care services, and other health and health related issues.

The sample is designed to meet and optimize two objectives:

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

The CHIS sample is representative of California's non-institutionalized population living in households.

This series of reports describes the methods used in collecting data for CHIS 2007, the fourth CHIS data collection cycle, which was conducted between June 2007 and early March 2008. The previous CHIS cycles (2001, 2003, and 2005) are described in similar series, available at http://www.chis.ucla.edu/methods.html.

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. The data are widely 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.

1.2 Sample Design Objectives

To achieve the sample design objectives stated above, CHIS employed a multi-stage sample design. For the first time, the random-digit-dial (RDD) sample included telephone numbers assigned to both landline and cellular service. For the landline RDD sample, the state was divided into 44 geographic sampling strata, including 41 single-county strata and three multi-county strata comprised of the 17 remaining counties. Within each geographic stratum, residential telephone numbers were selected, and within each household, one adult (age 18 and over) respondent was randomly selected. In those households with adolescents (ages 12-17) and/or children (under age 12), one adolescent and one child were randomly selected; the adolescent was interviewed directly, and the adult most knowledgeable about the child's health completed the child interview.

Table 1-1 shows the 44 sampling strata for CHIS 2007, which include 41 independent county strata. A sufficient number of adult interviews were allocated to each stratum to support the first sample design objective—to provide health estimates for adults at the local level. The geographic stratification of the state was the same as that used in CHIS 2005. In the first two CHIS cycles there were 41 total sampling strata, including 33 individual counties. The CHIS 2007 samples in Los Angeles and San Diego Counties were enhanced with additional funding by implementing further stratification within county.

The main landline RDD CHIS sample size is sufficient to accomplish the second objective. 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.

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 2007, the goal was to complete 800 interviews statewide with adults in cell-only households. Because data are not available for numbers assigned to cellular service to support the same level of geographic stratification as the landline sample, the cell RDD sample was stratified by area code. Sampled cellular numbers were screened to identify whether they belonged to cell-only households. Cellular numbers from households with landline telephone numbers were considered out of scope. If the sampled number was shared by two or more adult members of a cell-only household, one household member was selected for the adult interview. Otherwise, the

adult owner of the sampled number was selected. No interviews with adolescents or about children were conducted from the CHIS 2007 cell RDD sample.

Table 1-1. California county and county group strata used in the CHIS 2007 sample design

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

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

In an attempt to assess nonresponse bias, CHIS 2007 also included an area probability sample in Los Angeles County, with a target of 800 completed adult interviews. A clustered sample was selected from US Postal Service address lists, stratified by Los Angeles County Service Planning Area (SPA). Within each SPA, a number of smaller geographic areas (*segments* composed of blocks or groups of blocks) were selected, and within each segment specific addresses were selected. Sampled addresses for which a telephone number could be matched were initially treated the same as landline RDD cases, except that adolescent and child interviews were not attempted. Matched addresses where a screening interview could not be completed by telephone and all unmatched addresses were then assigned to recruiters who visited the sampled addresses in person to attempt to obtain cooperation.

1.3 Data Collection

To capture the rich diversity of the California population, interviews were conducted in five languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, and

Korean. These languages were chosen based on analysis of 2000 Census data to identify the languages that would cover the largest number of Californians in the CHIS sample that either did not speak English or did not speak English well enough to otherwise participate.

Westat, a private firm that specializes in statistical research and large-scale sample surveys, conducted the CHIS 2007 data collection under contract with the UCLA Center for Health Policy Research. For the landline RDD sample, Westat staff interviewed one randomly selected adult in each sampled household, and sampled one adolescent and one child if present in the household and the sampled adult was the parent or legal guardian. Up to three interviews could have been completed in each household. In households with children where the sampled adult was not the screener respondent, 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 new for CHIS 2005 and substantially increased the yield of child interviews. While numerous subsequent attempts were made to complete the adult interview, there were completed child and/or adolescent interviews in households for which an adult interview was not completed. For the cell RDD and area samples, only one randomly selected adult in each household was interviewed. Table 1-2 shows the number of completed adult, child, and adolescent interviews in CHIS 2007 by the type of sample (landline RDD, surname list, cell RDD, and area sample).

Table 1-2. Number of completed CHIS 2007 interviews by type of sample and instrument

Type of sample	Adult	Child	Adolescent
Total all samples	51,048	9,913	3,638
Landline RDD	48,791	9,818	3,622
Surname list	451	95	16
Cell RDD	825	N/A	N/A
Area (Los Angeles County)	981	N/A	N/A

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Interviews in all languages were administered using Westat's computer-assisted telephone interviewing (CATI) system. The average adult interview took about 35 minutes to complete. The average child and adolescent interviews took about 17.5 minutes and 20 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 8 percent of the adult interviews were completed in a language other

than English, as were almost 16 percent of all child (parent proxy) interviews and 7 percent of all adolescent interviews.

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

1.4 Response Rates

The overall response rate for CHIS 2007 is a composite of the screener completion rate (i.e., success in introducing the survey to a household and randomly selecting an adult to be interviewed) and the extended interview completion rate (i.e., success in getting one or more selected persons to complete the extended interview). To maximize the response rate, especially at the screener stage, an advance letter in five languages was mailed to all sampled telephone numbers for which an address could be obtained from reverse directory services. An advance letter was mailed for approximately 67 percent of the sampled telephone numbers. As in CHIS 2005, a $2 bill was included with the advance letter to promote cooperation.

The CHIS 2007 screener completion rate for the landline sample was 35.5 percent, and was higher for households that were sent the advance letter. For the cell phone sample, the screener completion rate was 30.5 percent in cell-only households. For the area sample, the screener response rate was 32.0 percent, compared with 31.5 percent for the landline sample in Los Angeles County. The extended interview completion rate for the landline sample varied across the adult (52.8 percent), child (73.7 percent) and adolescent (44.1 percent) interviews. The adolescent rate includes getting permission from a parent or guardian. The adult interview completion rate for the cell sample was 52.0 percent, and for the area sample 69.0 percent. Multiplying the screener and extended rates gives an overall response rate for each type of interview. The percentage of households completing one or more of the extended interviews (adult, child, and/or adolescent) is a useful summary of the overall performance of the landline sample. For CHIS 2007, the landline sample household response rate was 21.1 percent (the product of the screener response rate and the completion rate at the household level of 57.9 percent). All of the household and person level response rates vary by sampling stratum. For more information about the CHIS 2007 response rates, please see *CHIS 2007 Methodology Series: Report 4 – Response Rates*.

Table 1-3. CHIS 2007 survey topic areas by instrument

Health status	**Adult**	**Teen**	**Child**
General health status, height and weight	✓	✓	✓
Days missed from school due to health problems		✓	
Health conditions	**Adult**	**Teen**	**Child**
Asthma	✓	✓	✓
Diabetes	✓	✓	
Gestational diabetes	✓		
Heart disease, high blood pressure	✓		
Infertility	✓		
Falls (elderly)	✓		
Attention deficit disorder (ADD/ADHD), developmental disorders			✓
Parental concerns with child development			✓
Mental health	**Adult**	**Teen**	**Child**
Mental health status	✓	✓	✓
Perceived need, use of mental health services	✓	✓	✓
Emotional functioning	✓	✓	✓
Health behaviors	**Adult**	**Teen**	**Child**
Dietary intake	✓	✓	✓
Physical activity and exercise	✓	✓	✓
Sedentary time		✓	
Parental influence over diet and exercise			✓
Parental exposure to messages about obesity, smoking			✓
Developmental screening tests			✓
Colon cancer screening	✓		
Flu Shot	✓	✓	✓
Alcohol and tobacco use	✓	✓	
Drug use		✓	
Sexual behavior, STD testing	✓	✓	
Birth control practices		✓	
Women's health	**Adult**	**Teen**	**Child**
Pap test screening, mammography screening, hormone replacement therapy	✓		
Emergency contraception	✓	✓	
HPV – knowledge and awareness; vaccine use and attitudes	✓	✓	
Pregnancy status	✓	✓	
Dental health	**Adult**	**Teen**	**Child**
Last dental visit		✓	✓
Not getting needed care		✓	✓
Days missed from school due to dental problems		✓	✓
Dental insurance coverage	✓	✓	✓

Table 1-3. CHIS 2007 survey topic areas by instrument (Continued)

Food insecurity/hunger	**Adult**	**Teen**	**Child**
Availability of food in household over past 12 months	✓		
Access to and use of health care	**Adult**	**Teen**	**Child**
Usual source of care, visits to medical doctor	✓	✓	✓
Emergency room visits	✓	✓	✓
Delays in getting care (prescriptions, tests, treatment)	✓	✓	✓
Communication problems with doctor	✓		✓
Ability to understand medical instructions	✓		
Health insurance	**Adult**	**Teen**	**Child**
Current insurance coverage, spouse's coverage, who pays for coverage	✓	✓	✓
Health plan enrollment, characteristics of plan	✓	✓	✓
Whether employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months	✓	✓	✓
Reasons for lack of insurance	✓	✓	✓
Public program eligibility	**Adult**	**Teen**	**Child**
Household poverty level	✓		
Program participation (TANF, CalWorks, Public Housing, Food Stamps, SSI, SSDI, WIC)	✓	✓	✓
Assets, alimony/child support/social security/pension	✓		
Eligible for Medi-Cal and healthy families	✓	✓	✓
Reason for Medi-Cal nonparticipation among potential eligibles	✓	✓	✓
Neighborhood	**Adult**	**Teen**	**Child**
Neighborhood safety, use of parks	✓	✓	✓
Mode of local transportation	✓		
Interpersonal Violence	**Adult**	**Teen**	**Child**
Experiencing violence from intimate partner, details of most recent experience	✓	✓	
Experiencing violence from acquaintance	✓	✓	
Parental involvement/adult supervision	**Adult**	**Teen**	**Child**
Adult presence after school		✓	
Child's activities with family			✓

Table 1-3. CHIS 2007 survey topic areas by instrument (Continued)

Child care and school attendance	Adult	Teen	Child
Current child care arrangements			✓
Paid child care	✓		
Preschool/school attendance, name of school		✓	✓
Employment	**Adult**	**Teen**	**Child**
Employment status, spouse's employment status	✓		
Work in last week	✓		
Hours worked at all jobs	✓		
Income	**Adult**	**Teen**	**Child**
Respondent's and spouse's earnings last month before taxes	✓		
Household income (annual before taxes)	✓		
Number of persons supported by household income	✓		
Respondent characteristics	**Adult**	**Teen**	**Child**
Age, gender, height, weight, education	✓	✓	✓
Race and ethnicity	✓	✓	✓
Marital status	✓		
Sexual orientation	✓	✓	
Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home, English language proficiency	✓	✓	✓

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

The CHIS response rate is comparable to response rates of other scientific telephone surveys in California, such as the 2007 California Behavioral Risk Factor Surveillance System (BRFSS) Survey. Using calculations that are as comparable as possible to those of CHIS 2007, the combined screener and adult response rate for the 2007 BRFSS is 18.7 percent, exactly the same as that for the CHIS 2007 landline sample. California as a whole and the state's urban areas in particular are among the most difficult parts of the nation in which to conduct telephone interviews. Survey response rates tend to be lower in California than nationally, and over the past decade response rates have been declining both nationally and in California. Information about CHIS data quality and nonresponse bias is available at http://www.chis.ucla.edu/dataquality.html.

Adults who completed at least approximately 80 percent of the questionnaire (i.e., through Section K (on employment, income, poverty status, and food security), after all follow-up attempts were exhausted to complete the full questionnaire, were counted as "complete." At least some items in the employment and income series or public program eligibility and food insecurity series are missing from those cases that did not complete the entire interview.

Proxy interviews were allowed for frail and ill persons over the age of 65 who were unable to complete the extended adult interview in order to avoid biases for health estimates of elderly persons that might otherwise result. Eligible selected persons were recontacted and offered a proxy option. For 168 elderly 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. (Note: questions not administered in proxy interviews are given a value of "-2" in the data files.)

1.5 Weighting the Sample

To produce population estimates from the 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 2007 accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons;
- Reduce biases occurring because nonrespondents may have different characteristics than respondents;
- Adjust, to the extent possible, for undercoverage in the sampling frames and in the conduct of the survey; and
- Reduce the variance of the estimates by using auxiliary information.

As part of the weighting process, a household weight was created for all households that completed the screener interview. This household weight is the product of the "base weight" (the inverse of the probability of selection of the telephone number) and a variety of adjustment factors. The household weight is used to compute a person-level weight, which includes adjustments for the within-household sampling of persons and nonresponse. The final step is to adjust the person-level weight using a raking method so that the CHIS estimates are consistent with population control totals. Raking is an iterative procedure that forces the CHIS weights to sum to known population control totals from an independent data source (see below). The procedure requires iteration to make sure all the control totals, or raking dimensions, are simultaneously satisfied within a specified tolerance.

Population control totals of the number of persons by age, race, and sex at the stratum level for CHIS 2007 were created primarily from the California Department of Finance's 2007 Population Estimates and 2007 Population Projections. The raking procedure used 11 raking dimensions, which are combinations of demographic variables (age, sex, race, and ethnicity), geographic variables (county, Service Planning Area in Los Angeles County, and Health Region in San Diego County), household composition (presence of children and adolescents in the household), and socio-economic variables (home ownership and education). The socio-economic variables are included to reduce biases associated with excluding households without landline telephones from the sample frame. One limitation of using Department of Finance 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). 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.

1.6 Imputation Methods

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

Two different imputation procedures were used by Westat to fill in missing responses for items essential for weighting the data. The first imputation technique was a completely random selection from the observed distribution of respondents. This method was used only for a few variables when the percentage of the items missing was very small. The second technique was hot deck imputation without replacement. The hot deck approach is probably the most commonly used method 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 to a survey item form a pool of donors, while the nonrespondents are 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.

Hot deck imputation was used to impute the same items in CHIS 2003, CHIS 2005 and CHIS 2007 (i.e., race, ethnicity, home ownership, and education).

UCLA-CHPR imputed missing values for nearly every variable in the data files other than those handled by Westat and some sensitive variables in which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2007 were low, with most variables missing valid responses for less than 2% of the sample. However, there were a few exceptions where item nonresponse rate was greater than 20%, such as household income.

The imputation process conducted by UCLA-CHPR started with data editing, sometimes referred to as logical or relational imputation: for any missing value, a valid replacement value was sought based on known values of other variables of the same respondent or other sample(s) from the same household. For the remaining missing values, hierarchical sequential hot-deck imputation with 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 set of control variables. The control variables were ranked in order from the most to the least important. This procedure allowed control variables to be dropped if certain conditions (such as the minimum number of donors) were not met. The control variables were dropped sequentially, starting from the variable ranked least important. Once a responding case was used as a donor, it was dropped from the donor pool preventing using one donor multiple times.

Control variables used in forming donor pools for hot-decking always included the following: gender, age group, race/ethnicity, poverty level (based on household income), educational attainment, and region. Other control variables were also used depending on the nature of the imputed variable. Among the control variables, gender, age, race/ethnicity and regions were imputed by Westat. UCLA-CHPR then imputed household income and educational attainment in order to impute other variables. Household income, for example, was imputed using the hot-deck method within ranges from a set of auxiliary variables such as income range and/or poverty level.

The imputation order of the other variables followed the questionnaire. After all imputation was done, logic checks and edits were performed once again to ensure consistency between the imputed and nonimputed values on a case-by-case basis.

1.7 Methodology Report Series

A series of five methodology reports is available with more detail about the methods used in CHIS 2007:

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

For further information on CHIS data and the methods used in the survey, visit the California Health Interview Survey Web site at http://www.chis.ucla.edu or contact CHIS at CHIS@ucla.edu.

2. WEIGHTING ADJUSTMENTS CHIS 2007 SAMPLE WEIGHTS

This chapter introduces the concept of weighting and provides some background on the weights developed for analyzing CHIS 2007 survey data. Weighting is a process that attempts to make the estimates from the survey respondents representative of the total population that was sampled by accounting for the chances of selecting units into the sample and making adjustments for imperfections in the frame. The process begins with a base weight that is then adjusted to account for additional stages of sampling, nonresponse and undercoverage.

As described in *CHIS 2007 Methodology Series: Report 1 - Sample Design*, CHIS 2007 has samples from four different kinds of frames; landline RDD (including geographically-based supplements), surname list, cellular, and area. Three sets of weights were produced for data analysis, representing different combinations of these samples: (1) one set for the landline RDD and surname list samples; (2) another set for all telephones samples that include the landline, surname, and cell samples; and (3) a third for all of the telephone samples plus the area sample combined. Although this chapter deals with the sets of weights and their adjustments, it begins with the general reasons why fully adjusted weights should be used. It also describes the details, advantages, and disadvantages of weighting. Note that only the third set of weights is included in the final data as the first two are created for internal methodological assessments.

2.1 Weighting Approach

The philosophy used in CHIS 2007 weighting is a standard design-based multiple-frame methodology that is consistent with the sampling methods used. The dual-frame approach has been used in previous cycles of CHIS to combine and weight the landline and surname list samples. The same multiple frame approach can be used with some adjustments to produce weights that add the new cell and area samples.

The procedures used in CHIS are consistent across all sets of weights for all users and analyses in CHIS. Using the same analytic methods in a unified procedure also makes it much simpler for analysts to examine characteristics for many issues, such as analyzing estimates from the main and landline supplemental samples for San Diego[1]. Operationally, the weighting steps are

[1] This is a landline geographic supplemental sample.

similar and can be applied at the same time across samples (whenever appropriate), streamlining the weighting process and reducing the time required to produce the sets of weights.

Weights are applied to CHIS 2007 sample data to estimate aggregate statistics at the state and county levels. In particular, sample weighting was carried out to accomplish the following objectives:

- Compensate for differential probabilities of selection and sampling rates for households and persons;
- Reduce biases occurring because nonrespondents may have different characteristics from respondents;
- Adjust, to the extent possible, for undercoverage in the sampling frames and in the conduct of the survey; and
- Reduce the variance of the estimates by using auxiliary information.

The sets of weights developed for CHIS 2007 and their use in analysis are listed in Table 2-1. These samples and weights provide an opportunity to asses the effect of nonsampling errors due to nonresponse in a telephone survey and noncoverage of households without landline telephones on CHIS estimates. The results of these analyses are important and their findings will affect future cycles of the survey.

The first set of weights, the combined landline/list weights, was created to produce estimates that are consistent with previous cycles of CHIS. The second set of weights, the combined landline/list/cell telephone sample weights, was created to produce estimates that avoid the coverage bias of a landline sample that excludes cell-only households. These estimates should be more accurate for the entire California population than those from the landline/list sample only. The third set of weights, the area sample, covers only Los Angeles County and its primary use is to evaluate potential bias due to both noncoverage and nonresponse. The last set of weights combines the four samples. This set of weights includes the largest number of completed interviews for analyses. Details of the creation of the combined weights are in Appendix C.

As in previous cycles of CHIS, the sets of weights include a single weight created for each adult, child, and adolescent completed interview in the samples.

Table 2-1. Set of weights for CHIS 2007

	Weight set name	Samples	Use/analysis	Geographic area	Records to weight
1	Landline	Landline Surname list	▪ Statewide and county population estimates for trending with prior cycles. ▪ Nonresponse and coverage bias when only a landline sample is used in future cycles.	State	Adult, adolescent, child
2	Landline – Cell phone	Landline Surname list Cell	▪ Statewide and county population estimates. ▪ Coverage bias of landline due to cell-only households.	State	Adult
3	Area	Area	▪ Telephone nonresponse analysis in Los Angeles County ▪ Coverage bias of the landline sample in Los Angeles County due to households without a landline.	LA County	Adult
4	Combined	Landline Surname list Cell Area	▪ Statewide and county population estimates for trending with prior cycles.	State	Adult, adolescent, child

2.2 Weighting Adjustments

The final weight for a completed CHIS interview is the product of a series of sequential adjustments. The starting point is the development of a base weight defined as the inverse of the probability of selection from the stratum frame. The household weight is created adjusting the base weight for

- Subsampling screener refusals for conversion attempt;
- Unknown residential status;
- Supplemental list sample eligibility;
- Screener interview nonresponse;
- Cell-only household adjustment; and
- Multiple telephone numbers.

These adjustments are described in Chapter 3.

The household weight is then adjusted to create a person weight for each type of extended interview. For the adult weights, the following factors are included:

- Probability of selection of the adult;
- Extended adult interview nonresponse adjustment; and
- Raking adjustment to person-level control totals.

The child and adolescent weights are more complex because of the method used to sample children (see *CHIS 2007 Methodology Series: Report 1 - Sample Design*). For these weights, the adjustment factors include:

- Use of child-first procedure for households in which adolescents and children are sampled at the end of the screener interview and may be interviewed before the adult interview;
- Section G adult extended interview nonresponse adjustment for households in which adolescents and children are sampled at the end of Section G of the adult interview;
- Probability of selection of the child or adolescent;
- Extended child and adolescent interview nonresponse adjustment; and
- Raking adjustment to person-level control totals.

The expressions for the weighting factors and adjustments for the person weights are given in Chapters 4, 5, and 6. The derivation of the population control totals is described in Chapter 7. The imputation process and the imputed variables to support the weighting process are described in Chapter 8. Chapter 9 discusses the methods for variance estimation for CHIS 2007.

Appendix A contains tables showing the frame and sample sizes. Appendix B contains tables that show the effect of each step of the weighting process at the household and person levels. Throughout this report, we refer to specific tables and rows in Appendix B that indicate how the weights were adjusted.

2.3 Nonresponse adjustments

In an ideal survey, all the units in the inference population are in the sample frame and all those in the sample participate in the survey. In practice, neither of these conditions occurs.

Some units are not included in the frame (undercoverage) and some of the sampled units do not respond (nonresponse). If undercoverage and nonresponse are not addressed, then estimates from the survey may be biased. In CHIS 2007, the weights of those who respond are adjusted to represent undercovered persons in the population and nonrespondents in the sample. The approaches used to account for these two sources of missing data begin with adjusting for nonresponse.

Nonresponse results in biases in survey estimates when the characteristics of respondents differ from those of nonrespondents. The size of the bias depends on the magnitude of this difference and the response rate (see Groves, 1989). The purpose of adjusting for nonresponse is to reduce the bias. A weighting class adjustment (see Brick and Kalton, 1996) method is the type of nonresponse adjustment procedure used in CHIS 2007. In this procedure, nonresponse adjustment weights are computed and applied separately by cell, where a cell is defined using characteristics known for both nonrespondents and respondents. For example, the county associated with each telephone number is known, even if there are some misclassifications in the assignment. Thus, county can be used to define cells, and weighting adjustments can be computed separately for each of these cells. The more similar either response patterns or survey characteristics are within the cells, the larger the bias reduction in the adjustment.

The drawback to nonresponse adjustment is that it increases the variability of the weights and increases the sampling variance of the estimates (Kish, 1992). A nonresponse adjustment is beneficial only when the reduction in bias more than compensates for the increase in variance. When the cells contain sufficient cases and the adjustment factors do not become inordinately large, the effect on variances is often modest. Large adjustment factors usually occur in cells with small numbers of respondents. To avoid this situation, cells with few cases are "collapsed" or combined to form a new cell with a larger number of cases.

The operational objective for nonresponse adjustment in CHIS 2007 was to define adjustment cells for which response rates vary considerably and to avoid cells with either a small number of cases or a large adjustment factor. Since county-level estimates are important, the county was nearly always included in the definition of the cells. Oh and Scheuren (1983) discuss some of the statistical features associated with making these adjustments.

As noted above, nonresponse adjustment classes can be formed only if data are available for both responding and nonresponding units. Since the nonresponse adjustment is done for each stage of data collection, the data available for forming cells are different for each stage.

For screening interviews, the nonresponse unit is a household (or more accurately a telephone number), and data must be available for all households. For extended interviews, the nonresponse adjustment is done by type of person (adult, child, or adolescent). At this level, data from the screening interview can be used to define cells.

The approach to adjusting for undercoverage is somewhat different from that for nonresponse because noncovered units or persons were never eligible to be sampled. The undercoverage adjustment procedure uses data from external sources (control totals) in a process called poststratification (Holt and Smith, 1979). The primary objective of poststratification is to dampen potential biases arising from a combination of response errors, sampling frame undercoverage, and nonresponse. A secondary objective is to reduce sampling errors, which is important because CHIS 2007 sample sizes within counties are fairly modest for some subclasses. In general, the sample is poststratified to as many independent figures as possible, subject to some constraints. In this discussion we use the term poststratification loosely and intend it to include raking, a form of multidimensional poststratification (see Brackstone and Rao, 1979). In CHIS 2007, the control totals are mainly derived from the 2007 California Department of Finance Population Estimates and 2007 California Department of Finance Provisional Population Estimates (State of California, Department of Finance, 2006a, 2006b), the 2004 American Community Survey (U.S. Census Bureau, 2003), and the Census 2000 Summary File 1 for California published by the U.S. Census Bureau (U.S. Census Bureau, 2001). Details of the creation of the control totals at the person level are described in Chapter 7.

2.4 Combining Samples

In this section, we describe how the samples were combined to create the sets of weights for CHIS 2007. Before explaining the approach for combining the samples, we examine the relationship between the different frames and samples.

Consider the different samples as illustrated in Figure 2-1, which shows the relationship for the Los Angeles County stratum.

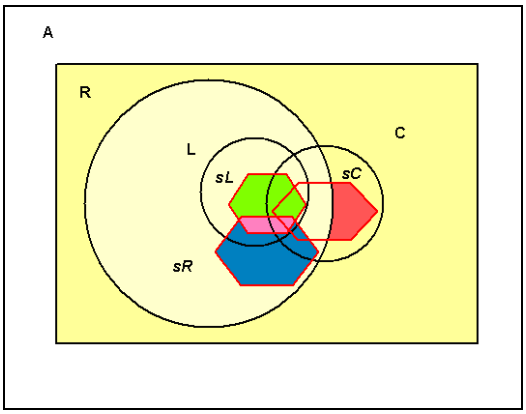


Figure 2-1. Landline, list, cell phone and area frames, and samples in CHIS 2007

Note: The figure is not drawn to scale. The sizes of the samples relative to the frames are smaller than shown in the figure.

Let A be all eligible households in Los Angeles County (represented by the large yellow rectangles in the diagrams). Let R be the landline frame (the large circle in the diagrams) that includes all the households that have a landline telephone and L be the surname list frame (small circle enclosed within R). Note that by definition, R is included within A and that L is included within R (i.e., $L \subset R$). Let C be the frame of the cell phone households that includes all households with no landline but with one or more cell phones, (i.e., $C \cap \overline{R}$), and households with both types of telephone service (i.e., $C \cap R$). Notice that the cell frame, C, is not encompassed by R, but crosses both R and A. Let s_A, s_R, s_L, and s_C be the area, landline, surname list, and cell phone samples respectively.

The cell phone sample s_C is used to screen for cell-only households. The screening process implicitly creates two nonoverlapping sampling strata. The first stratum includes sampled households with cell phone only, (i.e., $C \cap \overline{R}$) and the second stratum includes sampled households with both telephone services (i.e., $C \cap R$). Households in the second stratum are not eligible for the cell phone survey. Thus, there is no need to compute the multiple probability of selection for households in the second stratum. Cell-only households that were classified into the first stratum have only one chance to be selected in the sample, so their base weights are computed as the inverse of the probability of selection. This method works well when respondents can be classified into strata during screening. However, there may be some undercoverage bias and losses in efficiency due to misclassification, as sometimes happens with screening respondents.

Operationally, since the cell-only sample is a nonoverlapping stratum, the cell-only sample is weighted at the same time as the landline sample applying the appropriate weighting

adjustments (e.g., the adjustment for subsampling for refusal conversion is only applicable to the landline sample).

Consider now the list sample s_L and the landline sample s_R. By definition, the surname frame *L* is contained in the landline frame *R,* so all telephones on the surname frame have two probabilities of selection (one for landline sample and the other for surname sample). Since the screening process creates a separate stratum for the cell-only sample, the inclusion of the cell-phone sample does not affect the probability of selection of households in the landline and surname samples. In other words, the landline and surname samples can be weighted following the same methods used in previous cycles of CHIS. Base weights depend on whether or not the telephone number is found on the surname frame (the information on about whether any of the landline sample cases are on the surname frames is available from the surname list vendor). With this process the landline and surname samples can be combined. Hence, when we refer to *R* and S_R in the remainder of this section we implicitly include *L* and S_L.

Finally, consider the area sample s_A, an approximate self-weighting two-stage sample of addresses. Clusters of census blocks were sampled at the first stage, followed by sampling of addresses from the selected clusters. For purposes here, s_A is treated as an independent sample. That is, the base weights for the area sample were calculated as if the addresses had only a single chance at selection. See *CHIS 2007 Methodology Series: Report 1 - Sample Design* for more details on the sample selection.

The expression of the base weights is described in more detail in Section 3.1. To create the combined landline/list/cell phone weight, the weights of samples are pooled together and raked to the state population totals in the last step of weighting.

3. HOUSEHOLD WEIGHTING

For all CHIS samples, the first step in the weighting process for CHIS 2007 is creating a household weight for each completed screener interview. The household weight is not used for analytical purposes because the only data captured at the household level in the screener interview are for sampling purposes. However, the household weight is a key element for the computation of the person weights (i.e., adult, child, and adolescent).

This chapter is divided into nine sections, each describing the steps involved in creating the household weights. The first section reviews the creation of base weights. Subsequent sections describe the adjustments made to the base weights. These adjustments account for refusal conversion subsampling, unknown eligibility residential status, supplemental list sample eligibility, unknown presence of children in the household, screener nonresponse, households with multiple telephone numbers, and subsampling in child supplemental samples.

Knowledge of the sampling methods used in CHIS 2007 is essential to understanding the weighting procedures. We assume anyone interested in the weighting procedures is already familiar with the sampling approach – details are in *CHIS 2007 Methodology Series: Report 1 - Sample Design*.

3.1 Base Weights

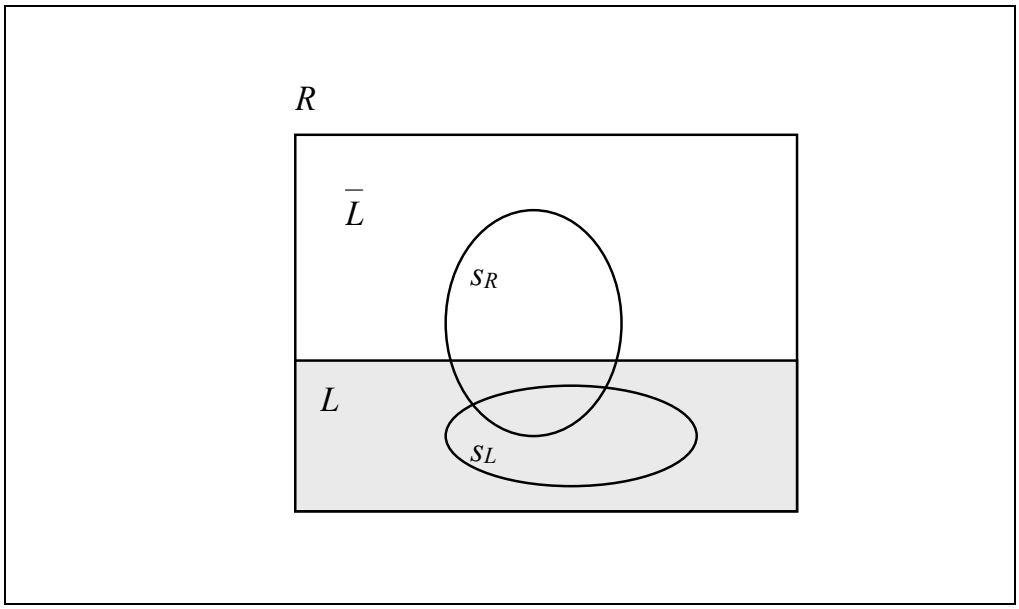
A base weight is created for each sampling unit in the different CHIS samples. For the landline, list, and cell samples, the sampling unit is the telephone number; for the area sample, the sampling unit is the address.

3.1.1 Landline and Surname List Base Weight

Base weights for the landline/list sample are computed as the inverse of the probability of selection of the telephone number. In CHIS 2007, telephone numbers were drawn from the landline frame and five mutually exclusive surname list frames (Korean, Vietnamese,

Korean-Vietnamese[2], Vietnamese and any other race but Korean, and Korean and any other race but Vietnamese). The base weights reflect the multiple probability of selection of telephone numbers between the landline and list frames.

Figure 3-1 shows the relationship between the landline frame and a single surname list frame for a single sampling stratum. The figure also shows the relationship between the samples drawn from each frame. In order to create the household base weights, we consider all landline telephone households in California as either being on the list (L) or as only being eligible for sampling from the landline sample ($\overline{L}$) as shown in Figure 3-1. The relationships are discussed in detail below.



* The figure is not drawn to scale. The sizes of the list frame (L) and list and landline samples (s_L and s_R) are smaller than shown in the figure.

Figure 3-1. Relationship between the landline frame (R), landline sample (s_R), list frame (L), and list sample (s_L) for a single stratum

The notation in the figure follows:

R	the landline frame containing all telephone numbers;
L	the list frame (i.e., surnames);
$\overline{L}$	all telephone numbers not found on the list – we assume that all the numbers in the list are found in R, and $R = L \cup \overline{L}$;
s_R	the simple random sample drawn from the frame R; and

[2] A separate frame was created for surnames that could not be readily identified as of Korean or Vietnamese origin. Because this frame contains surnames that are associated with both ethnic groups, this frame is referred to as the Korean-Vietnamese surname list.

s_L the simple random sample drawn from the frame *L*.

Define the following:

N_R the number of telephone numbers in the frame R;
N_L the number of telephone numbers in the frame L;
n_R the sample size (number of telephone numbers) of s_R; and
n_L the sample size (number of telephone numbers) of s_L.

Notice that the landline sample s_R can be separated into two parts: s_{RL}, the portion of s_R that is found in the list (L) and $s_{R\bar{L}}$, the portion of s_R that is not found in the list $(\bar{L})$. The sample sizes for each portion are n_{RL} and $n_{R\bar{L}}$, respectively, and $n_R = n_{RL} + n_{R\bar{L}}$.

Consider L and $\bar{L}$ as two separate strata within the frame R. Since s_R is a simple random sample within R, the sample $s_{R\bar{L}}$ can be viewed as a simple random sample of size $n_{R\bar{L}}$ drawn from the $N_{\bar{L}}$ elements from stratum $\bar{L}$. Similarly, the sample s_{RL} can be viewed as a simple random sample of size n_{RL} drawn from the N_L elements from stratum L. In stratum L, there is a second sample s_L (the list sample). Since both samples s_L and s_{RL} are simple random samples, we can view them as a single sample of size $n_{RL} + n_L$ drawn from the N_L elements from stratum L. Notice that s_{RL} and s_L are not necessarily mutually exclusive; i.e., s_{RL} and s_L may contain duplicate telephone numbers. These numbers were removed from s_L during the sample selection.

The landline and list base weights can be expressed as follows:

- For sampled records that could only be sampled from the landline frame (landline numbers not found in the list L):

$$HHBW_{\bar{L}i} = \frac{N_{\bar{L}}}{n_{R\bar{L}}};$$

- For sampled records from the list and sampled records from the landline frame that are found in the list L (duplicate telephone numbers were eliminated from the list):

$$HHBW_{Li} = \frac{N_L}{n_{RL} + n_L}.$$

Creating these weights required classification of every telephone number by whether or not it was on the list irrespective of how it was sampled. It is easy to show that the resulting weights are composite weights derived by averaging the landline and list samples using a composite factor proportional to the sample sizes. Thus, this base weight produces an unbiased estimate in the traditional design-based framework.

The total telephone numbers in the landline frame and list frames (N_R and N_L) are computed separately. The landline sample was drawn using an RDD list-assisted approach from a stratified frame of 100 banks[3] with at least one listed telephone number in the state of California. Using this approach, a bank is drawn from the frame and two digits are randomly generated to complete the sampled telephone number. Therefore, the total number of telephone numbers in the landline frame in stratum *h*, N_{Rh}, is computed as

$$N_{Rh} = 100 \cdot NBANKS_h ,$$

where $NBANKS_h$ is the number of 1+ banks in March of 2007 in stratum *h*. A "1+" bank is defined as a 100 bank with at least one working telephone number.

Records on the list frames were assigned to landline sampling strata by linking telephone exchanges to the counties in the same way as for the landline sample. The list size by stratum (N_{Lh}) is the number of records in the list assigned to stratum *h*.

As described in *CHIS 2007 Methodology Series: Report 1 - Sample Design,* the landline sample was drawn from strata defined as counties or groups of counties except for Los Angeles, San Diego, Orange, and Santa Clara. In Los Angeles County, 13 subsampling strata were created by the combination of areas with high/low concentration of Koreans and Vietnamese and eight Special Planning Areas (SPAs). In San Diego County, eight substrata were created by the combination of six health regions in combination with high/low concentration of Koreans and Vietnamese. Two substrata based on the concentration of Koreans and Vietnamese were created for Orange and Santa Clara Counties. The definition of the sampling strata and substrata, in addition to the number of telephone numbers in the landline frame, the number of sample cases, and base weights by frame type (landline, Korean only, Vietnamese only, Korean or Vietnamese only , Vietnamese and another group, and Korean, and another group lists), is shown in Appendix A,

[3] A bank is defined as 100 consecutive telephone numbers with the same first eight digits including area code.

Tables A-1 and A-2. Table B-1 in Appendix B (rows 1.1 through 1.3) list the sample counts, sums of base weights, and coefficients of variation by sampling stratum for these samples.

3.1.2 Cell Phone Base Weight

The cell phone sample was drawn for a stratified random sample of 1,000-series blocks dedicated to wireless service (NXXTYPE types 04, 55, 60) or PCS (personal communication service types 65, 68)[4]. The cell sample base weight is similar to the landline weight except that there are 1,000 numbers in each bank. The sampling stratum was defined by the area code of telephone numbers assigned to wireless service. For more details on the cell phone sample design, see *CHIS 2007 Methodology Series: Report 1 - Sample Design.*

Let $CPBW_{hi}$ be the base weight for the *i*-th sampled cell phone number in the *h*-th stratum (defined by area codes), the base weight is computed as

$$CPBW_{hi} = \frac{NC_h}{n_h},$$

where n_h is the total sampled numbers in stratum h, and NC_h is the total numbers in stratum h, computed as $NC_h = 1000 \cdot NS_h$ where NS_h is the number of 1,000 blocks in stratum *h*. Note that the stratum definition for the cell phone sample is different from that of the landline sample. The definition of sampling strata, in addition to the number of telephone numbers in the frame, number of sampled cases, and average base weights are shown in Appendix A, Table A-3.

3.1.3 Area Sample Base Weight

The area sample used an approximately self-weighting two-stage sample of addresses from Los Angeles County. The first stage was a stratified (by Service Planning Area or SPA), probability-proportional-to-size sample of clusters of census blocks. The second stage was the sampling of addresses from the selected clusters. For more details, see *CHIS 2007 Methodology Series: Report 1 - Sample Design.*

[4] There are some additional technical restrictions in the sampling, such as making sure the number can be dialed into and that toll-free numbers are excluded.

For the area sample, the household base weight is product of the inverse of probability of selection of the sampling units at each stage. The expression of the base weight of a household *i* in cluster *c* in stratum *h* in the area sample is

$$ABW_{hci} = \frac{1}{PC_{hc}PA_{hci}},$$

where PC_{hc} is the probability of selection of the cluster *c* computed as $PC_{hc} = n_h C_{hc} / \sum_{c=1}^{C_h} C_{hc}$, where n_h is the number of clusters sampled in stratum *h*, C_{hc} is the measure of size of cluster *c* (i.e., number of households in the cluster based on the 2000 census file), and $\sum_{c=1}^{C_h} C_{hc}$ is the total measure of size in stratum h. The variable PA_{hci} is the probability of selection of address *i* in cluster *c* computed as $PA_{hci} = m_{hc} / A_{hc}$, where m_{hc} is the number of addresses to sample in the cluster *c* and A_{hc} is the total number of addresses in cluster *c*.

The definition of sampling strata, in addition to the number of clusters and households in the frame, the number of sampled cases, and average base weights are shown in Appendix A, Table A-4.

The expression of the base weight $HHBSW_{hi}$ for the household *i* in stratum *h* for the area samples is

$$HHBSW_{hi} = \begin{cases} HHBW_{hi} & \text{if } i \in \text{ landline, list or geographic samples} \\ CPBW_{hi} & \text{if } i \in \text{ cell phone sample} \\ ABW_{hi} & \text{if } i \in \text{ area sample} \end{cases},$$

where $HHBW_{hi}$, $CPBW_{hi}$, and ABW_{hi} are the landline/list, cell phone, and area sample base weights, respectively. Because the weights are adjusted within stratum, the subscript *h* will be omitted in the description of the weights in the following sections as a notational convenience.

3.2 Refusal Subsampling Adjustment

The base weights were adjusted to reflect the differential refusal conversion efforts made during data collection. Subsampling of refusals has been used since CHIS 2003 for increasing the efficiency of data collection (Brick et al., 2005). The rationale for refusal subsampling depends on two observations: refusal cases comprise the majority of screener nonresponse in CHIS (as in most telephone surveys); and substantial effort is expended to gain cooperation in households in which a member refuses to participate in the study at the screener level. The subsampling of refusals shifts some resources from the less productive, labor-intensive task of refusal conversion to the more productive task of completing extended interviews. The subsampling rates are computed trying to balance the data collection costs and precision of the estimates. The principles of subsampling for refusal conversion are well established (see Hansen and Hurwitz, 1946, and Elliott, Little, and Lewitzky, 2000). The method is used in other surveys, including the American Community Survey conducted by the U.S. Census Bureau.

There are some disadvantages of this method. First, a larger sample of telephone numbers than would otherwise be selected is required. A weighting adjustment is required to account for the subsampling. Those cases that refuse and are subsampled are weighted to represent themselves and the cases that refuse and are not subsampled. This weighting decreases the precision of the survey estimates, but only very slightly. Another disadvantage is that response rates should be weighted. Unweighted rates should not be used to assess response patterns because they do not reflect the subsampling of refusal conversion cases, and unweighted response rates in CHIS cannot be compared with other surveys.

Subsampling for refusal conversion in CHIS 2007 was implemented only at the screener level in the landline sample. There was no subsampling of extended interview refusals in any sample or type of interview (i.e., adult, child, or adolescent). Each number in the landline sample was randomly assigned with two flags during sampling. When refusals were encountered at the screening stage of data collection for a landline sample case, only numbers with the first flag were eligible for first refusal conversion follow-up. Among respondents who refused for the first time, only those with the second flag were eligible for second refusal conversion. The numbers subsampled for first refusal follow-up are generally fielded first so that any refusal cases can be worked completely. This procedure differs somewhat from subsampling refusals in previous cycles of CHIS where cases subsampled for conversion were all subject to both first and second refusal conversion attempts.

The first flag was pre-assigned to approximately 80 percent of the telephone numbers from the landline sample. These telephone numbers were eligible for first refusal conversion. The second flag was pre-assigned to approximately two-thirds of the telephone numbers. These telephone numbers were eligible for full conversion (i.e., first and second refusal conversion).

Before adjusting the weights for screener interview refusal subsampling, telephone numbers were classified into screener refusal groups using their refusal status (i.e., whether the respondent ever refused) and the value of the refusal conversion flag as shown in Table 3-1.

Table 3-1. Screener refusal groups for landline sample

Screener refusal group	Respondent ever refused screener interview?	First Refusal Subsampling Flag	Second Refusal Subsampling Flag	Description
NRef	No	N/A	N/A	Households where respondent did not refuse the screener interview (includes complete and incomplete screener interviews)
RefC1	Yes	Yes	No	Households where respondent refused the screener interview and only first refusal conversion procedures were used
RefC2	Yes	Yes	Yes	Households where respondent refused the screener interview and both first and second refusal conversion procedures were used
RefNC	Yes	No	No	Households where respondent refused the screener interview and refusal conversion procedures were not used

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

The first refusal subsampling adjusted weight, $HHA1W_i$, is:

$$HHA1W_i = HHA1F_i * HHBSW_i$$

where $HHA1F_i$ is the first refusal subsampling adjustment factor computed as:

$$HHA1F_c = \begin{cases} \dfrac{\sum_{i\in(RefC1,RefNC)} HHBSW_i\delta_i(c)}{\sum_{i\in RefC1} HHBSW_i\delta_i(c)} & \text{If } i \in RefC1 \\ 0 & \text{If } i \in RefNC \\ 1 & \text{If } i \in N\text{Re}f \end{cases},$$

where the groups $RefC1$, $RefNC$, and $NRef$ are defined in Table 3-1, $HHBSW_i$ is the base weight, and $\delta_i(c)$ is 1 if the number is in sampling stratum c and is zero otherwise.

The second refusal subsampling adjusted weight, $HHA2W_i$, is:

$$HHA2W_i = HHA2F_i * HHA1W_i$$

where $HHA2F_i$ is the second refusal subsampling adjustment factor computed as:

$$HHA2F_c = \begin{cases} \dfrac{\sum_{i\in(RefC2,\text{Re}fC1)} HHA1W_i\delta_i(c)}{\sum_{i\in RefC2} HHA1W_i\delta_i(c)} & \text{If } i \in RefC2 \\ 0 & \text{If } i \in RefNC \\ 1 & \text{If } i \in N\text{Re}f \end{cases},$$

where the groups $RefC2$, $RefNC$, and $NRef$ are defined in Table 3-1.

Towards the end of the field period, additional telephone numbers were released in selected strata to meet the targets for the number of completed interviews. Refusal conversion was not attempted for these additional releases so these are treated as being not sampled for conversion. Table B-1 in Appendix B (rows 2.1 through 3.5) shows the sum of the weights before and after the refusal conversion subsampling adjustments.

This adjustment did not affect cases from the surname samples because all the numbers from these samples were included in refusal conversion. This adjustment was not applicable to households in the cell phone and area sample, so the adjustment factor was set to one for those samples.

3.3 Subsampling Cases with a Mailable Address

In the CHIS 2001 the efficiency of the sample was improved by stratifying the telephone numbers by mailable status and subsampling the strata at different rates (Brick, et al., 2002). After CHIS 2003, the effectiveness of oversampling mailable numbers was examined. It was determined that the expected gain in efficiency would be modest and might not be sufficient to justify the additional complexity associated with this procedure. As a result, mailable status subsampling was not implemented in 2005 and 2007.

3.4 Unknown Residential Status Adjustment

Telephone numbers with unknown residential status are those that could not be classified by residential status at the end of data collection despite being dialed many times. They are telephone numbers answered only by answering machines or some combination of answering machine and ring no answer (screener disposition code of *NM*) or never answered even by a machine (screener disposition of *NA*). Prior to adjusting the weights to account for telephone numbers with unknown residential status, the number of eligible residential telephone numbers among those numbers with unknown residential status was estimated. This estimate was also used in the computation of the response rates described in *CHIS 2007 Methodology Series: Report 4 - Response Rates*.

In CHIS 2007, the estimated proportion of unknown residential telephone numbers considered residential (p_{res}) was computed separately for the landline, surname, cell phone, and area samples. The proportion p_{res} was computed using the CASRO (1982) recommendation as the proportion of the resolved or observed sample units that are residential (or occupied units for the area sample). Since telephone numbers were sampled with different selection probabilities and were adjusted differentially for refusal conversion, the weighted number of telephone numbers were used rather than the number of cases (unweighted) to compute p_{res}. This use of weights also compensates for the under- and oversampling implemented in different geographic areas.

Table 3-2 shows the values of p_{res} for the landline sample. As expected, the estimated proportion of residential households is much lower for answering machines coded as "possible nonresidential" compared to those coded as "possible residential." For example, the estimated proportion of residential households in urban strata with mailable addresses and answering machines coded as possible residential is 91.8 percent, while the estimated proportion of

those coded as not urban mailable- possible nonresidential is 20.2 percent. The lowest percentages of residential telephone numbers are for the numbers that were not mailable and were never answered ("no machine") or had answering machine messages coded as possible nonresidential.

Table 3-2. Estimated residential proportion for the landline sample

Mail status	Urban status	Answering machine code	p_{res}
Mailable	Urban	No machine	0.677
Mailable	Urban	Possible residential	0.918
Mailable	Urban	Possible nonresidential	0.166
Mailable	Urban	Unknown	0.851
Mailable	Not urban	No machine	0.717
Mailable	Not urban	Possible residential	0.915
Mailable	Not urban	Possible nonresidential	0.202
Mailable	Not urban	Unknown	0.856
Not mailable	Urban	No machine	0.036
Not mailable	Urban	Possible residential	0.862
Not mailable	Urban	Possible nonresidential	0.065
Not mailable	Urban	Unknown	0.527
Not mailable	Not urban	No machine	0.038
Not mailable	Not urban	Possible residential	0.858
Not mailable	Not urban	Possible nonresidential	0.069
Not mailable	Not urban	Unknown	0.538

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 3-3 shows the values of p_{res} for the list samples. Since there were no differences by type of list sample, the values of p_{res} were computed combining the cases from the lists.

Table 3-3. Estimated residential proportion for the list samples

Answering machine code	p_{res}
No machine	0.938
Answering machine possible residential	0.951
Answering machine possible business	0.393
Answering machine unknown	0.308

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 3-4 shows the values of p_{res} for the cell phone sample. This proportion was computed by type of phone (ported or wireless assigned) and region. There were differences in the estimated proportions by these groups; the value of p_{res} for ported cell phones is on average 18 percentage points higher than the p_{res} for wireless assigned cell phones.

Table 3-4. Estimated residential proportion for the cell phone samples by type of cell phone and region

Type of cell phone	Region	p_{res}
Ported	1 - Northern & Sierra Counties	0.616
	2 - Greater Bay Area	0.914
	3 - Sacramento Area	0.869
	4 - San Joaquin Valley	0.675
	5 - Central Coast	0.677
	6 - Los Angeles	0.797
	7 - Other Southern California	0.843
Wireless assigned	1 - Northern & Sierra Counties	0.485
	2 - Greater Bay Area	0.610
	3 - Sacramento Area	0.657
	4 - San Joaquin Valley	0.562
	5 - Central Coast	0.562
	6 - Los Angeles	0.637
	7 - Other Southern California	0.626

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 3-5 shows the values of p_{res} for the area sample by service planning area. The observed proportion of households that are residences in the area sample is much larger than that in the telephone samples.

Table 3-5. Estimated residential proportion for the area sample by Service Planning Area

Service Planning Area	p_{res}
Antelope Valley	0.962
San Fernando	0.978
San Gabriel	0.970
Metro	0.963
West	0.958
South	0.942
East	0.975
South Bay	0.976

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

The estimated proportion of residential households among the unknown residential telephone numbers or addresses p_{res} is then used to adjust the weights for unknown residential status. The residential status adjusted weight, $HHA3W_i$, is

$$HHA3W_i = HHA3F_i * HHA2W_i,$$

where $HHA3F_i$ is the unknown residential status adjustment factor computed as:

$$HHA3F_i = \begin{cases} \dfrac{\sum_{i \in RES} HHA2W_i + \sum_{i \in UNK\_RES} p_{res} * HHA2W_i}{\sum_{i \in RES} HHA2W_i} & \text{If } i \in RES \\ 0 & \text{If } i \in UNK\_RES \end{cases},$$

where the group *RES* denotes telephone numbers identified as residential and *UNK_RES* denotes telephone numbers with unknown residential status. In the area sample, the group *RES* denotes occupied units and *UNK_RES* denotes units with unknown occupancy status.

This adjustment is done separately by sample type. In the landline sample, the adjustment is done within sampling stratum by mailable status. In the list sample, the adjustment is by surname list type. In the area sample, it is by SPA. Although this adjustment was not applied to the cell phone sample (see next section), but the information shown in Table 3-4 shows that there were differences between the ported and wireless assigned cell phones. The adjustment factor $HHA3F_i$ was set to one for all records in this sample. Table B-1 in Appendix B (rows 4.1 through 4.5) shows the sum of weights before and after making the adjustment for unknown residential status for the landline and surname samples and Table B-3 (rows 2.1 through 2.5) for the area sample. As expected, the adjustment is smaller for the area sample because it is easier to determine residential status in the field.

3.5 Sample Eligibility Nonresponse Adjustment

After adjusting the weights for unknown residential status, the weights are adjusted for eligibility in the samples where screening is used to identify eligible respondents. Screening is used to identify households with adults of Korean or Vietnamese descent in the surname samples and adults in cell-only households in the cell phone sample. Therefore, this adjustment is only applicable to these two samples.

The supplemental surname list samples (i.e., Korean only, Vietnamese only, Korean and other, Vietnamese and other, and Korean-Vietnamese list samples) were screened to identify adults of Korean and Vietnamese descent. The weights for these samples were adjusted to account for households in which the ethnic origin of the adults (i.e., Korean or Vietnamese) could not be

determined. Telephone numbers from the surname list samples were eligible only if one or more adults in the household considered themselves of Vietnamese or Korean descent[5].

Households with at least one adult from one of these ethnic groups are referred to as "list-eligible" households. If a household from the supplemental sample was found to be list-eligible, then one adult from these ethnic groups was selected for the extended interview. If the household was not list-eligible (i.e., no adults of Vietnamese or Korean descent), then the screener interview was terminated and the case was coded as a list-ineligible.

Screening on eligibility and retaining only list-eligible households in the surname supplemental samples was a simple method for increasing the number of Korean and Vietnamese extended interviews in CHIS 2007 and previous cycles. The information on the ethnic origin of the adults was used to eliminate unnecessary interviews of adults with different ethnic origin, who were eligible in the landline sample.

Household list eligibility could not be determined for nonresponding households in the surname list samples, and the weights had to be adjusted for unknown list eligibility. The weights of the households with unknown list eligibility were distributed between the list-eligible and ineligible households in the surname samples. The assumption in this adjustment was that the proportion of list-eligible/ineligible households among the households with unknown list eligibility was the same as the observed proportion in the sample with known eligibility. The cases were classified in response groups as indicated in Table 3-6.

Table 3-6. List eligibility response groups

List eligibility response status group		Description
L_E	List-eligible	Household from the surname list sample with at least one list-eligible adult (i.e., adult of Korean or Vietnamese descent).
L_IN	List-ineligible	Household from the surname list sample without any list-eligible adult (i.e., no adults of Korean or Vietnamese descent).
L_UNK	List eligibility unknown	Household from the surname list sample where the eligibility of the adults could not be determined.
L_NA	List eligibility not screened	Household from all other samples (not screened for eligible ethnicity).

[5] Question SC6A1 of the screener interview asked, "Do any of these adults who live in your household consider themselves to be Korean or Vietnamese or of Korean or Vietnamese descent?"

The list eligibility nonresponse adjusted household weight, $HHA4W_i$, is computed as

$$HHA4W_i = HHA4F_c * HHA3W_i,$$

where $HHA4F_c$ is the list eligibility nonresponse adjustment factor computed as

$$HHA4F_c = \begin{cases} \dfrac{\sum_{i\in(L\_E,\,L\_IN,\;\;L\_UNK)} HHA3W_i\delta_i(c)}{\sum_{i\in(L\_ER,L\_IN)} HHA3W_i\delta_i(c)} & i\in(L\_E, L\_IN) \\ 0 & i\in L\_UNK \\ 1 & i\in L\_NA \end{cases},$$

where the groups *L_E*, *L_IN*, *L_UNK*, and *L_NA* are defined in Table 3-4, and $\delta_i(c)$ is 1 if the number is in list eligibility nonresponse adjustment cell *c* and is zero otherwise. The nonresponse adjustment cells correspond to the list sample type (i.e., Korean, Vietnamese, Korean-Vietnamese samples). Table B-1 in Appendix B (rows 5.1 through 5.5) shows the sum of weights before and after the list eligibility nonresponse adjustment.

Screening is also used to identify cell-only households (i.e., with and without landline service) in the cell phone sample. Cell phone households with a landline were identified by asking questions about telephone usage in households in the screener questionnaire. These questions are used to develop a weighting adjustment for the cell-only households.

Telephone numbers in the cell phone sample are eligible for an extended interview if the contacted cell phone is for personal use, if the phone is used by an adult, and the household does not have landline telephone service. If these conditions are met, then one adult is sampled among those adults who share the cell phone. If there are no eligible adults who use the phone, then the screener interview is terminated and the case is coded as an ineligible household for the cell phone sample. If there are adults but the household has landline service, then questions were asked about the cell phone usage, but no adult was sampled for an adult interview.

We adjust the weights to account for numbers in which the "cell phone" status cannot be ascertained. Before adjusting the weights for cell phone status, the cases are classified in response groups based on cell status as indicated in Table 3-7.

Table 3-7. Cell phone eligibility response groups

Cell phone response status group		Description
C_E	Cell eligible	Household with a cell phone service
C_IN	Cell ineligible	Not a cell phone number, no adults in household, business use only, out of area, teen cell phone
C_UNK	Unknown cell eligibility	Refusal, language problem

The value of $HHA4F_c$ for the cell phone sample is computed the same way as in the surname sample by replacing the groups *L_E*, *L_IN*, and *L_UNK* by the groups *C_E*, *C_IN*, and *C_UNK* as defined in Table 3-7. The nonresponse adjustment cells are the cell phone sampling strata.

3.6 Screener Nonresponse Adjustment

In this step, the household weight is adjusted to account for households that did not complete the screener interview. The nonresponse adjustment cells were created separately for the main landline and surname list samples and utilized information on the presence of children in the household from the screener.

In the first step of screener nonresponse adjustment we adjusted the weights to account for the presence of children in the household. The weights of nonresponding households with a known child status were distributed to responding households. This weight, $HHA5W_i$, is:

$$HHA5W_i = HHA5F_c * HHA4W_i,$$

where $HHA5F_c$ is the unknown presence of children adjustment factor computed as

$$HHA5F_c = \begin{cases} \dfrac{\sum_{i\in(SC\_KCS,\ SC\_UCS)} HHA4W_i\delta_i(c)}{\sum_{i\in SC\_KCS} HHA4W_i\delta_i(c)} & i \in SC\_KCS, \\ 0 & i \in SC\_UCS \end{cases},$$

where the group *SC_KCS* is the set of screener respondents with known child presence status, and *SC_UCS* is the set of screener nonrespondents with unknown child status, and $\delta_i(c)$ is 1 if the number is in screener nonresponse adjustment cell *c* and is zero otherwise.

In the second step of screener nonresponse adjustment we adjusted the weights to account for screener nonresponse among households with a known presence of children. This weight, $HHA6W_i$, is:

$$HHA6W_i = HHA6F_c * HHA5W_i,$$

where $HHA6F_c$ is the screener nonresponse adjustment factor computed as

$$HHA6F_c = \begin{cases} \dfrac{\sum_{i \in (SC\_R,\ SC\_NR)} HHA5W_i \delta_i(c)}{\sum_{i \in SC\_R} HHA5W_i \delta_i(c)} & i \in SC\_R, \\ 0 & i \in SC\_NR \end{cases},$$

where the group *SC_R* is the set of screener respondents, and *SC_NR* is the set of screener nonrespondents, and $\delta_i(c)$ is 1 if the number is in screener nonresponse adjustment cell *c* and is zero otherwise.

List-ineligible households (i.e., households with no adults of Korean or Vietnamese origin) from the surname list samples (group *R_IN* defined in the previous section) were considered as screener nonrespondents (group *SC_NR*) in this adjustment. Although these cases were households with only list-ineligible adults, they still represented households with eligible adults for the landline sample extended interview who were screened out. Table B-1 in Appendix B (rows 6.1 through 6.5) gives the sum of weights before and after the first screener nonresponse adjustment.

This adjustment was also applied to the area sample where addresses already confirmed as residential households are adjusted for those households who refused the screener interview. Table B-3 in Appendix B (rows 3.1 through 3.5) shows the sum of weights after and before these adjustments.

On the other hand, this adjustment was not applied to cases from the cell phone sample because this form of nonresponse was included in the previous step. The adjustment factors are set to one for the cell sample.

3.7 Cell-only household adjustment

In this next step, the cell phone household weights are benchmarked to the proportion of cell-only households in the population. In the cell sample, the weighted proportion of screener interviews reported as cell-only households prior to this adjustment was 39 percent. The estimate of cell-phone-only households among households that have cell phone service for the last 6 months of 2007 for the West region is 18 percent as reported in the National Health Interview Survey (NHIS) (Blumberg and Luke, 2008). The reason for the difference in the estimate from CHIS and the NHIS estimate is most likely due to the differential response rate of cell phone users; cell-only users are much more likely to answer their cell phone than are less-frequent cell phone users.

In order to benchmark the weights, households from the cell phone sample were classified in response groups shown in Table 3-8 based the cell phone usage. The cell phone usage was created using the answers to the screener questions CINTRO_3 or CINTRO_4 (whether there is a landline in the household) and C2 (relative use of cell and landlines) for those with completed screeners.

Table 3-8. Cell phone response groups

Response group	Landline in household?	Cell phone use	Description	Proportion (p_C)
C_only	No	Only	Cell-only households	0.18
C_high	Yes	High	Households with both landline and cell phone service where the respondent receives all or almost all calls on a cell phone	0.17
C_med	Yes	Medium	Households with both landline and cell phone service where the respondent receives some calls on a cell phone	0.36
C_low	Yes	Low	Households with both landline and cell phone service where the respondent receives very few or no calls on cell phone	0.28

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

At this stage, the cell phone sample weights are adjusted so the proportion of households by phone use matched the NHIS figures shown in Table 3-8. The cell phone benchmarked weight for cell phone use, $HHA7W_i$, is computed as

$$HHA7W_i = HHA7F * HHA6W_i,$$

where $HHA7F$ is the list eligibility nonresponse adjustment factor. For the cell-only cases (group *C_only*), $HHA7F$ is computed as

$$HHA7F = p_C \cdot \frac{\sum_{i \in (C\_only, C\_high, C\_med\_C\_low)} HHA6W_i}{\sum_{i \in C\_only} HHA6W_i},$$

where p_C is the proportion from Table 3-6. Although similar expressions were used to adjust the groups *C_high, C_med,* and *C_low*, only the cell-only households are kept in the file for the following weighting adjustment.. This set of adjustments was not applied to the landline, list and area samples, and the adjustment factors were set to one for all records in these samples.

3.8 Multiple Telephone Adjustment

At the end of the screener interview, information about the existence of additional telephone numbers and their use in the household was collected. If more than one telephone number is used for residential purposes (not solely for business, fax or computer use, etc.), the household has a greater probability of selection because it could have been selected through any of the additional telephone numbers in the household. In such cases, the household weight is adjusted to reflect the increased probability of selection. The multiple telephone adjusted household weight, $HHA8W_i$, is computed as:

$$HHA8W_i = HHA8F_i * HHA7W_i,$$

where $HHA8F_i$ is the multiple telephone adjustment factor computed as:

$$HHA8F_i = \begin{cases} 0.5 & \text{If household } i \text{ has more than one residential telephone number} \\ 1 & \text{Otherwise} \end{cases}.$$

In this adjustment, we assume that there is at most one additional residential-use telephone number in the household. In a few cases, the same household was reached through two different telephone numbers found in the sample. For these cases, only one interview was conducted and the pair of telephone numbers was identified based on the information from the respondent. One telephone number was coded as duplicate (no interview associated with this number), and the weight of the other telephone number (with the completed interview) was not reduced by the multiple telephone factor because the household was sampled twice. Table B-1 in Appendix B (rows 7.1 through 7.5) identifies the sum of weights before and after the multiple telephone adjustment. This adjustment was not applied to the cell or area samples and therefore their adjustment factors were set to 1 in this step.

3.9 Household Poststratification

Since the inception of the survey, the household weights have been poststratified to household control totals. The poststratification adjustment used cells for households with and without a person under 18 years old by sampling stratum. The same control totals derived from the Census 2000 files for California have been used because no other source for the total number by these type households is available by individual counties. However, after the experience gained from producing weights for the combined 2005 landline sample and the 2005 pilot cell sample, it was decided not to poststratify the household weights in CHIS 2007.

The weights were adjusted for the differential response by type of household through raking dimension 10 at the last step in weighting. These dimensions control for the total number of adults in households. To do the poststratification at the household level would require combining the different samples at this step, adding complexity to the person level adjustments.

4. ADULT WEIGHTING

An adult final weight was created for each adult who completed the adult extended interview.[6] The initial adult weight is the product of the final household weight and the reciprocal of the probability of selecting the adult from all adults in the household for the landline and area samples. For the cell phone sample, the initial weight is the product of the final household weight and the number of adults in the household where the cell phone is shared; if the cell phone is not shared, the initial adult weight equals the final household weight. In subsequent steps, the initial adult weight is adjusted for nonresponse and raked to known control totals. For the landline sample (including the surname list sample) weights, undercoverage of adults that could not be interviewed because they reside in households without a landline telephone was compensated for by a raking adjustment that included a dimension to reduce the undercoverage bias. Details on creating the adult weights follow.

4.1 Adult Initial Weight

As described in *CHIS 2007 Methodology Series: Report 1 - Sample Design*, one adult was sampled with equal probability from all adults in the household using the Rizzo method (see Rizzo et. al., 2004). The initial adult weight is the product of the final household weight and the inverse of the probability of selection of the adult. The expression for the adult initial weight, $ADA0W_j$, is

$$ADA0W_j = ADCNT_i \cdot HHA8W_i ,$$

where $ADCNT_i$ is the total number of adults in household *i* for the landline and area samples and the number of adults in the household (if there are adults that share the cell phone) for the cell phone sample, and $HHA8W_i$ is the multiple telephone adjusted weight described in the previous chapter.

This sampling scheme for the cell phone sample assumes that, in cell-only households with more than one adult, each adult has a cell phone (or shares a different cell phone) if the sampled cell phone is not shared. If the cell phone is shared, it assumes that all adults in the household shared the same phone.

[6] Adult extended interviews are considered complete provided the adult completed through Section K on employment and income.

Appendix Tables B-4, B-7, and B-8 (rows 1.1 through 1.3) show the number of adults, sum of initial weights, and coefficient of variation for the landline/list sample for the state and by sampling stratum for the landline/list, cell phone, and area samples.

4.2 Adult Nonresponse Adjustment

Regardless of the sample (i.e., landline, cell phone, or area samples), some households completed the screener interview but the sampled adult did not complete the extended adult interview. In addition, in a few cases it was discovered during the extended interview that the sampled person was under 18 years of age and hence ineligible. To account for both sampled adults who did not complete the extended interview and for ineligible sampled persons, the adult initial weight was adjusted for extended interview nonresponse. Prior to making the adjustment, extended interviews were classified into response groups as indicated in Table 4-1.

Table 4-1. Extended interview response groups

Response status group		Description
ER	Eligible respondent	Adult who completed the extended interview
IN	Ineligible	Ineligible person
UNK	Unknown eligibility	Sampled adult could not be contacted and eligibility verified for extended interview

The adult nonresponse adjusted weight, $ADA1W_i$, is

$$ADA1W_i = ADA1F_c \cdot ADA0W_i,$$

where $ADA1F_c$ is the adult nonresponse adjustment factor given by

$$ADA1F_c = \begin{cases} \dfrac{\sum_{i\in(ER,IN,UNK)} ADA0W_i \cdot \delta_i(c)}{\sum_{i\in(ER,IN)} ADA0W_i \cdot \delta_i(c)} & i \in (ER, IN) \\ 0 & i \in UNK \end{cases},$$

where *ER, ENR* and *IN* are defined in Table 4-1, *c* indicates the adult extended interview nonresponse adjustment cell, and $\delta_i(c) = 1$ if the adult belongs to cell *c* and is zero otherwise.

Table 4-2 lists the variables that were considered in defining the nonresponse adjustment cells. All of these have been examined in previous CHIS cycles expect for the adult screener respondent status. This variable was included in 2008 because adults are more likely to respond if they are both the screener and extended respondents. Because of the child-first procedure, a child and/or adolescent interview could be completed prior to the adult interview. In such households, the response propensity of the adult differed from that in households where the child-first procedure was not used.

A nonresponse analysis showed that sex, child-first interview status, age group, and adult screener respondent were the best variables for creating nonresponse cells. Nonresponse cells with fewer than 30 respondents or with large adjustment factors were combined with adjacent cells. All the cells were created within sampling stratum. Appendix B, Tables B-4, B-7, and B-8 (rows 2.1 through 2.5) show the sum of weights before and after the nonresponse adjustment for the landline/surname, cell phone, and area samples. Ineligible persons were dropped following this weighting step.

Table 4-2. Variables used for the creation of nonresponse adjustment cells for the adult weights

Variable	Levels	
Sex of adult respondent	1.	Male
	2.	Female
Child-first interview (landline and area sample only)	1.	Child-first procedures in affect
	2.	Child-first procedures not in affect
Adult age group	1.	18-30 years old
	2.	31-45 years old
	3.	46-65 years old
	4.	65 years or older
Adult screener Respondent	1.	Adult screened was a respondent
	2.	Adult screened was not a respondent

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

4.3 Adult Trimming Factors

Before benchmarking the adult weights to the known total of adults in California in 2007, we examined the distribution of the weights to determine if there were very large weights that could have a large effect on either the estimates or the variances of the estimates. When observations with large weights were found, the weights for these cases were reduced in a process called trimming.

As in previous cycles, we computed statistics to identify influential weights that were candidates for trimming. These statistics and other variations were studied in detail in Liu et al. (2004). The first statistic is a function of spacing of the weights. Let $w_{(1)},\ldots,w_{(n)}$ be the order statistics for the adult weights $w_1,\ldots,w_n$ and define "spacing" z_i as the distance (difference) between a ranked weight $w_{(i)}$ and the next ranked weight $w_{(i-1)}$ (i.e., $z_i = w_{(i)} - w_{(i-1)}$). The statistic $d5\_space_i$ for a ranked $w_{(i)}$ is defined as

$$d5\_space_i = \frac{z_i}{z_{i-1} + z_{i-2} + z_{i-3} + z_{i-4} + z_{i-5}}.$$

The second statistic used computes the distance between a weight and the next largest weight relative to the size of the weight. The statistic is

$$rel\_space_i = \frac{z_i}{w_{(i)}} \times 10.$$

We also computed a third statistic defined as

$$diff\_dist_i = distance_i - distance_{i-1},$$

where $distance_i$ is the relative distance for the weight $w_{(i)}$ computed as

$$distance_i = \frac{\left| w_{(i)} - median\,(\mathbf{w}) \right|}{MAD},$$

where $\mathbf{w} = (w_i,...,w_n)^t$ and the median absolute deviation $MAD = median\left(\left|w_i - median\,(\mathbf{w})\right|\right)$.

The three statistics for the largest 20 weights in each stratum were examined separately. When all three statistics were greater than 1 then the case was a primary candidate for trimming. The final decision on trimming involved the inspection of the weight distribution within sampling stratum.

The trimmed weight $TRMW_i$ is computed as

$$TRMW_i = TFACT_i * ADA1W_i,$$

where $TFACT_i$ is the trimming factor for the sampled adult *i* given by

$$TFACT_i = \begin{cases} 1 & \text{if the weight } i \text{ is not trimmed} \\ t_i & \text{otherwise} \end{cases}$$

where $0 < t_i < 1$.

The number of records that were trimmed[7] depended on the set of weights being created. In the combined landline and list weights, 78 records were trimmed where the trimming factor ranged between 0.4055 and 0.9983. In the combined landline, list and cell phone sample, 33 records were trimmed and the trimming factor ranged between 0.12884 and 0.81584. In the area sample weights, 17 records were trimmed and the trimming factor ranged between 0.5974 and 0.8718. Tables B-4, B-7, and B-8 (rows 3-1 to 3-3) show the strata with trimmed weights and the sum of weights before and after trimming for the different weights.

4.4 Adult Raked Weight

The next step in the adult weighting was raking the trimmed weights to population control totals to produce estimates consistent with the 2007 California Department of Finance (DOF) Population Estimates. Included in the raking adjustment is an undercoverage adjustment for adults in households without landline telephone discussed earlier. The specific control totals and the method used to create them are described in Chapter 7.

Raking is a commonly used estimation procedure in which estimates are controlled to marginal population totals. It can be thought of as a multidimensional poststratification procedure because the weights are poststratified to one set (a dimension) of control totals, then these adjusted weights are poststratified to another dimension. The procedure continues until all dimensions are adjusted. The process is then iterated until the control totals for all dimensions are simultaneously satisfied (at least within a specified tolerance). Raking is also described in more detail in Chapter 7.

The adult raked weight, $RAKEDW_i$, can be expressed as

$$RAKEDW_i = TRMW_i \cdot \prod_{k=1}^{K} RAKEDF_{k_l} \ ,$$

[7] The trimming was done prior to the raking adjustment; however, it was an iterative process. After the trimming and raking, the distribution of the weights was re-examined, and new decisions were made about trimming. This might have changed the decision about which weights should be trimmed or the magnitude of the trimming factor. If the decision was made, the trimmed and raked weights were discarded and new trimming and raking were undertaken. The number of trimmed weights reported here is at the completion of the overall process.

where $RAKEDF_{k_l}$ is the raking factor for dimension *k*, level *l* in which adult *i* is. For example, if the 4th dimension (*k* =4) is sex with two levels (*l*=1 for male and *l*=2 for female), then the raking factor for this dimension is $RAKEDF_{4_1}$ for the adult male. The raking factors are derived so the following relationship holds for every raking dimension *k*, and level *l*,

$$CNT_{k_l} = \sum_i \delta(k_l)_i \cdot RAKEDW_i \ ,$$

where CNT_{k_l} is the control total, and $\delta(k_l)_i = 1$ if the adult *i* is in level *l* of dimension *k* and zero otherwise.

The first set of weights, which include landline/list weights CHIS were produced by raking the landline and list sample weights. This set of weights is consistent with the weights from previous cycles of CHIS. Table B-4 (rows 3.3 and 4.2) shows sum of weights before and after the raking adjustment for the landline list combined sample.

The second set of weights, which includes the landline, list, cell-phone sample weights were created by raking to the same control totals used for the firs set of weights. However, only the adult weights were kept for analysis because the child and adolescent weights were almost identical to the first set of weights. Table B-7 (rows 3.3 and 4.2) shows the sum of weights of cell only adults in the combined landline, list and cell phone sample. Table B-8 (rows 3.3 and 4.2) shows the extended interview sums of weights for adults in the area sample.

5. CHILD WEIGHTING

A final child weight was created for all completed child extended interviews in the landline/list samples (no children or adolescents were selected in the cell phone or area samples). The steps for the child weighting are similar to those for adults described in the previous chapter. One exception is an additional weighting adjustment needed to account for nonresponse in a section of the adult interview where the majority of the children were sampled. A more complete discussion of this adjustment is found in Section 5.1. The format of this chapter follows that for the adult weighting, with the creation of the child initial weights and the adjustments for nonresponse, trimming, and finally raking.

5.1 Household-Level Adjustment

The main difference between the child (and adolescent) weighting procedures and those of the adults is that adults were always sampled in the screener. Children and adolescents could be selected at the end the screener interview or in Section G of the adult extended interview. The selection of children at the end of the screener interview is called child-first procedure. Consequently, weights for those children and adolescents must be further adjusted to account for nonresponse at the adult interview level. On the other hand, weights of the child-first children and adolescents were not adjusted for adult nonresponse.

Telephone numbers were classified into completion groups (SECGST) by Section G completion status and their child-first interview status as shown in Table 5-1.

Table 5-1. Section G completion groups

Section G completion group (SECGST)	Child-first interview?	Section G completed by adult?	Description
C1st	Yes	N/A	Households where the child-first interview procedures occurred
NC1stGC	No	Yes	Households where the child-first interview procedures did not occur and section G was completed
NC1stGNC	No	No	Households where the child-first interview procedures did not occur and section G was not completed

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

To account for adults who did not complete Section G of the adult interview (hence, no child or adolescent could be sampled), the household final weight $HHA8W_i$ was adjusted. We refer to this adjusted weight as the Section G adjusted household weight, $HHA9W_i$, and it is

$$HHA9W_i = HHA9F_c * HHA8W_i,$$

where

$$HHA9F_c = \begin{cases} \dfrac{\sum_{i\in(NC1stGC,NC1stGNC)} HHA8W_i\delta_i(c)}{\sum_{i\in NC1stGC} HHA8W_i\delta_i(c)} & \text{If } i \in NC1stGC \\ 0 & \text{If } i \in NC1stGNC \\ 1 & \text{If } i \in C1st \end{cases},$$

and where the section G completion groups *C1st*, *NC1stGC*, and *NC1stGNC* are defined in Table 5-1, c denotes the Section G nonresponse adjustment cell, and $\delta_i(c)=1$ if the adult belongs to cell c and is zero otherwise. Following this adjustment, the weights were positive for all households with sampled adults who completed section G and either completed, partially completed, or did not complete the adult interview[8]. Note that this adjustment can be considered as a household adjustment in addition to the household weight.

The Section G nonresponse adjustment cells were created within sampling strata using a combination of the mailable status (known address/mailed letter, unknown address) and the presence of children and/or adolescents. The information about the presence of children and adolescents is collected during the screener interview.

5.2 Child Initial Weight

The initial child weight is the product of the adjusted household weight and the probability of sampling the child within the household. The selection of the child was done in two steps. In the first step, one adult was randomly selected among all adults in the household. In the second step, one child was randomly selected among all the children associated with the sampled

[8] If the adult interview was not completed in a case that did not use the child-first procedure, no attempt was made to complete a child/teen interview.

adult (i.e., the sampled adult is the parent or legal guardian of the child). If the sampled adult did not have an associated child, then no child was sampled even if there were children present in the household. See *CHIS 2007 Methodology Series: Report 1 - Sample Design* for information on the within-household person selection process.

Since the selection of a child within a household depends on the relationships among children and adults within the household, these relationships were defined before sampling children. The probability of selection reflects the fact that the sampled child could have been selected through the spouse/partner of the sampled adult if both are the parents or legal guardians[9] of the sampled child. Accordingly, the initial child weight, $CHA0W_j$, is

$$CHA0W_j = \frac{1}{CHPROB_j} HHA9W_i$$

where $HHA9W_i$ is the section G adjusted weight, $CHPROB_j$ is the probability of selecting the j^{th} child associated with the i^{th} sampled adult and is relatively complex. If the sampled adult does not have a spouse/partner living in the household or if the spouse/partner of the sampled adult is not the parent or legal guardian of the sampled child, then

$$CHPROB_j = \frac{1}{ADLTCNT} \cdot \frac{SACHMOS_j}{\sum_j SACHMOS_j}$$

where $ADLTCNT$ is the number of adults in the household and $SACHMOS_j$ is the measure of size of child j. The measure of size for a child is discussed in detail in *CHIS 2007 Methodology Series: Report 1 - Sample Design*, but we note here that within the same household children age 0 to 5 years have a measure of size twice that of children age 6 to 11 years. If the sampled adult has a spouse/partner living in the household and the spouse/partner of the sampled adult is the parent or legal guardian of the sampled child then

$$CHPROB_j = \frac{1}{ADLTCNT} \left(\frac{SACHMOS_j}{\sum_j SACHMOS_j} + \frac{SACHMOS_j}{\sum_k SPCHMOS_k} \right)$$

[9] If the spouse/partner of the sampled adult is living in the household.

where $ADLTCNT$ and $SACHMOS_j$ are defined as before and $SPCHMOS_k$ the measure of size child *k* associated with the spouse/partner of the sample adult. The number of sampled children and sum of the initial weights are in Table B-5 (rows 1.1 through 1.3).

5.3 Other Child Weighting Adjustments

Adjustments were made to the child weights for extended interview nonresponse, trimming influential weights, and raking to control totals. The child nonresponse adjustment is the same as the adult nonresponse adjustment described in Section 4.2, except the adjustment cells are defined differently. We initially created child nonresponse adjustment cells using three variables: household mailable status, sex of child, and age group (0-3, 4-7, and 8-11 years old) within sampling stratum. Since a majority of these cells had fewer than 30 respondents, we collapsed cells to increase the number of respondents in each cell. To do this we inspected adjustment factors separately by mailable status, sex, and age group at the state level to determine the variables with the most variable response rates. Using these results, for two sampling strata mailable status, sex and age group were used, and for the rest of the sampling strata the cells were defined by sampling stratum, sex of child, and age group. Any cells still containing fewer than 30 respondents were collapsed over age group. The two strata with smallest sample sizes were collapsed across both sex and age group. Table B-5 (rows 2.1 through 2.3) shows the number of sample records and sum of weights before and after the nonresponse adjustments.

The next step was to identify and trim large child weights. The process used for trimming the adult weights was applied to the child weights. As a result of applying the procedures, we identified and trimmed a total of 60 child weights in CHIS 2007. The trimming factors range from 0.3871 to 0.8334511. Table B-5 (rows 3.1 through 3.3) shows the distribution of trimmed weights by self-reported stratum and the sum of the weights before and after applying the trimming factors.

The trimmed child weights were then raked to population control totals to produce estimates consistent with the California Department of Finance 2007 population estimates. See Chapter 7 for the specific controls used. The expression for the raking adjustment is the same as that for adult weights described in Section 4.4. Table B-5 (rows 4.1 through 4.5) shows the counts and sum of weights before and after the raking adjustments.

6. ADOLESCENT WEIGHTING

In CHIS 2007, adolescents were sampled and responded to the interview for themselves after parental permission to conduct the interview. In this section, we describe the creation of analytic weights for the adolescent interview. Like the child weighting adjustments, the adolescent weighting applies only to the landline/list sample since adolescents were not interviewed in the area and cell phone samples.

6.1 Initial Adolescent Weights

The procedures for creating the adolescent weights are the same as those for creating the child weights described in Chapter 5. As with the child weighting, the initial weights for the adolescents incorporate the probability of sampling the adult and the probability of sampling an adolescent among the adolescents associated with the sampled adult. The initial weight, $TNA0W_i$, is

$$TNA0W_i = \frac{1}{TNPROB_j} \cdot HHA9W_i$$

where $HHA9W_i$ is defined in Chapter 5, and $TNPROB_i$ is computed the same as $CHPROB_i$. However, the measure of size is unity for all adolescents regardless of their age. Table B-6 shows the number of sampled adolescents (row 1.1) and the sum of the initial adolescent weights (row 1.2).

6.2 Other Adolescent Weighting Adjustments

The adolescent initial weight was then adjusted for nonresponse the same way the adult and child initial weights were adjusted. Note that nonresponse for the adolescent interview includes failure to obtain permission for the interview, as well as failure to interview the adolescent once permission was obtained. Table B-6 in Appendix B shows the nonresponse adjusted adolescent weight. Initially the adolescent nonresponse adjustment cells were created using household mailable status, sex of the adolescent, and age group (12-14 and 15-17 years old) within sampling strata. We inspected response rates separately by the three variables at the state level. After reviewing these rates, we created cells using sampling stratum, mailable status, sex and age

group. Cells containing fewer than 30 respondents were collapsed across age group first and then across mailable status and sex if necessary.

After the nonresponse adjustments, 55 influential weights were identified and trimmed, with factors ranging from 0.3374 to 0.8479. Table B-6 (rows 3.1 through 3.3) gives the trimmed weights by self-reported stratum and the sum of the weights before and after applying the trimming factors to the adolescent weights

In the last steps, the adolescent weights were raked to California DOF 2007 Population Estimates. See Chapter 7 for details on the control totals. The expression for the raking adjustment is the same as in the raking of the adult weights and the child weights. Table B-6 (rows 3.3 and 4.2) show the sum of weights before and after raking.

7. RAKING AND CONTROL TOTALS

This chapter describes the raking procedure and the development of control totals for CHIS 2007. The first section gives a general overview of raking and why this procedure was used in this and previous cycles of CHIS. The second section describes the 11 dimensions used to rake the different set of weights. Eight of the dimensions are defined by demographic variables and two are defined by socio-economic variables. The 11th dimension was created to reduce the bias associated with households without a landline telephone. The third section describes how the control totals were derived from the California DOF files.

7.1 Raking Procedure

Raking is an adjustment procedure in which estimates are controlled to marginal population totals. The main advantage of raking over poststratification is that raking allows the use of more auxiliary information. A limitation in poststratification is that each unit falls into only one adjustment cell and the number of respondents in a cell could be too small. With raking, the cell size is based on the distribution of each raking dimension. For example, with poststratification, only some cross-classified age/race/sex categories could be used in the adjustments, whereas with raking the full cross-classification is not needed, and important geographic data such as county can be included as dimensions. Raking may be thought of as a multidimensional poststratification procedure because the weights are basically poststratified to one set (a dimension) of control totals, then these adjusted weights are poststratified to another dimension. After all dimensions are adjusted, the process is iterated until the control totals for all the dimensions are simultaneously satisfied within a specified tolerance. Raking was also used in previous cycles of CHIS. Below, we describe the procedure in more detail. Brackstone and Rao (1979); Deville and Särndal (1992); and Kalton and Flores Cervantes (2003) also describe raking.

For simplicity, consider two auxiliary variables (or dimensions) with C and D classes, respectively. If we cross-classify the two variables into $C*D$ cells and the sample counts in some cells are small, then it is likely that the poststratified estimates may be unstable unless the cells in the cross-tabulation are collapsed. With the 11 dimensions used in CHIS 2007, the potential collapsing would be very extensive.

An alternative approach is to rake the weights to the marginal totals of the variables. The raking-adjusted estimator is design-unbiased in large samples and is very efficient in reducing the variance of the estimates if the estimates in the cross-tabulation are consistent with a model that ignores the interactions between variables. Collapsing is sometimes required with raking, but it is not as extensive as with poststratification.

The raked weights can be written as $\tilde{w}_{cd,i} = w_{cd}\hat{\alpha}_c\hat{\beta}_d$, where w_{cd} is the pre-raked weight of an observation in cell (c, d) of the cross-tabulation, $\hat{\alpha}_c$ is the effect of the first variable, and $\hat{\beta}_d$ is the effect of the second variable. Note that in this formulation there is no interaction effect; the weights are determined by the marginal distributions of the control variables. As a result, the sample sizes of the marginal distributions are the important determinants of the stability of the weighting procedure, not the cells formed by the cross-classification of the variables. Deficient cells (cells with small sample sizes) are thus defined in terms of the sample sizes of the marginal distributions, not of the cross-classified cells.

7.2 Raking Dimensions

The 11 dimensions used in CHIS 2007 are shown in Table 7-1. The first eight dimensions in Table 7-1 were created by combining demographic variables (age, sex, race, and ethnicity) and different geographic areas (city, county, region or group of counties, and state). The 9th, 10th, and 11th dimensions use additional variables. The 11th dimension was specifically created to adjust the weights for households without a landline telephone. This dimension was also used to create the combined landline/list/cell sample weights. Section 7.3 describes this adjustment and the variables used to create the levels for this dimension. The raking dimensions for CHIS 2007 are similar to those used in previous CHIS cycles, which all also include 11 dimensions.

Table 7-1. Definitions of the dimensions used in raking

Dimension	Level	Description	Categories	
1	Stratum (collapsed where necessary)	Age groups (3) x Sex (2)	11	Under 12 years, male
			12	Under 12 years, female
			21	12 to 17 years, male
			22	12 to 17 years, female
			31	18 years or older, male
			32	18 years or older, female

See note at end of table.

Table 7-1. Definitions of the dimensions used in raking (continued)

Dimension	Level	Description	Categories	
2	Stratum (collapsed where necessary)	Age groups (9)	1	Under 6 years
			2	6 to 11 years
			3	12 to 17 years
			4	18 to 24 years
			5	25 to 29 years
			6	30 to 39 years
			7	40 to 49 years
			8	50 to 64 years
			9	65 years or older
3	State	Age groups (13) x Sex (2)	11	Under 4 years, male
			12	Under 4 years, female
			21	4 to 7 years, male
			22	4 to 7 years, female
			31	8 to 11 years, male
			32	8 to 11 years, female
			41	12 to 14 years, male
			42	12 to 14 years, female
			51	15 to 17 years, male
			52	15 to 17 years, female
			61	18 to 24 years, male
			62	18 to 24 years, female
			71	25 to 30 years, male
			72	25 to 30 years, female
			81	31 to 37 years, male
			82	31 to 37 years, female
			91	38 to 45 years, male
			92	38 to 45 years, female
			101	46 to 53 years, male
			102	46 to 53 years, female
			111	54 to 64 years, male
			112	54 to 64 years, female
			121	65 to 77 years, male
			122	65 to 77 years, female
			131	78 years or older, male
			132	78 years or older, female
4	SPAs in Los Angeles Co., HRs in San Diego Co., Remainder of CA	SPAs (8), HRs (6), Remainder of CA (1)	1	SPA 1 – Antelope Valley
			2	SPA 2 – San Fernando
			3	SPA 3 – San Gabriel
			4	SPA 4 – Metro
			5	SPA 5 – West
			6	SPA 6 – South
			7	SPA 7 – East
			8	SPA 8 – South Bay
			9	HR 1 – North Coastal
			10	HR 2 – North Central
			11	HR 3 – Central
			12	HR 4 – South
			13	HR 5 – East
			14	HR 6 – North Inland
			15	Remainder of CA

See note at end of table.

Table 7-1. Definitions of the dimensions used in raking (continued)

Dimension	Level	Description	Categories	
5	Region (collapsed where necessary)	Race/ethnicity (7)	1	Latino
			2	Non-Latino White
			3	Non-Latino African American
			4	Non-Latino American Indian
			5	Non-Latino Asian
			6	Non-Latino Native Hawaiian
			7	Non-Latino Two or more races
6	State	Gender (2) x Race/ethnicity (7) x Age groups (4) (collapsed where necessary)	111	Male, Latino, under 12 years
			112	Male, Latino, 12 to 17 years
			113	Male, Latino, 18 to 64 years
			114	Male, Latino, 65 years or older
			121	Male, Non-Latino White, under 12 years
			122	Male, Non-Latino White, 12 to 17 years
			123	Male, Non-Latino White, 18 to 64 years
			124	Male, Non-Latino White, 65 years or older
			131	Male, Non-Latino African American, under 12 years
			132	Male, Non-Latino African American, 12 to 17 years
			133	Male, Non-Latino African American, 18 to 64 years
			134	Male, Non-Latino African American, 65 years or older
			141	Male, Non-Latino American Indian, under 12 years
			142	Male, Non-Latino American Indian, 12 to 17 years
			143	Male, Non-Latino American Indian, 18 to 64 years
			144	Male, Non-Latino American Indian, 65 years or older
			151	Male, Non-Latino Asian, under 12 years
			152	Male, Non-Latino Asian, 12 to 17 years
			153	Male, Non-Latino Asian, 18 to 64 years
			154	Male, Non-Latino Asian, 65 years or older
			161	Male, Non-Latino Native Hawaiian, under 12 years
			162	Male, Non-Latino Native Hawaiian, 12 to 17 years
			163	Male, Non-Latino Native Hawaiian, 18 to 64 years
			164	Male, Non-Latino Native Hawaiian, 65 years or older
			171	Male, Non-Latino Two or more races, under 12 years
			172	Male, Non-Latino Two or more races, 12 to 17 years
			173	Male, Non-Latino Two or more races, 18 to 64 years
			174	Male, Non-Latino Two or more races, 65 years or older
			211	Female, Latino, under 12 years
			212	Female, Latino, 12 to 17 years
			213	Female, Latino, 18 to 64 years
			214	Female, Latino, 65 years or older
			221	Female, Non-Latino White, under 12 years
			222	Female, Non-Latino White, 12 to 17 years
			223	Female, Non-Latino White, 18 to 64 years
			224	Female, Non-Latino White, 65 years or older
			231	Female, Non-Latino African American, under 12 years
			232	Female, Non-Latino African American, 12 to 17 years

See note at end of table.

Table 7-1. Definitions of the dimensions used in raking (continued)

Dimension	Level	Description	Categories	
			233	Female, Non-Latino African American, 18 to 64 years
			234	Female, Non-Latino African American, 65 years or older
			241	Female, Non-Latino American Indian, under 12 years
			242	Female, Non-Latino American Indian, 12 to 17 years
			243	Female, Non-Latino American Indian, 18 to 64 years
			244	Female, Non-Latino American Indian, 65 years or older
			251	Female, Non-Latino Asian, under 12 years
			252	Female, Non-Latino Asian, 12 to 17 years
			253	Female, Non-Latino Asian, 18 to 64 years
			254	Female, Non-Latino Asian, 65 years or older
			261	Female, Non-Latino Native Hawaiian, under 12 years
			262	Female, Non-Latino Native Hawaiian, 12 to 17 years
			263	Female, Non-Latino Native Hawaiian, 18 to 64 years
			264	Female, Non-Latino Native Hawaiian, 65 years or older
			271	Female, Non-Latino Two or more races, under 12 years
			272	Female, Non-Latino Two or more races, 12 to 17 years
			273	Female, Non-Latino Two or more races, 18 to 64 years
			274	Female, Non-Latino Two or more races, 65 years or older
7	State	Asian groups (5) x Age groups (4) (collapsed where necessary)	11	Non-Latino Chinese only, under 12 years
			12	Non-Latino Chinese only, 12 to 17 years
			13	Non-Latino Chinese only, 18 to 64 years
			14	Non-Latino Chinese only, 65 years or older
			21	Non-Latino Korean only, under 12 years
			22	Non-Latino Korean only, 12 to 17 years
			23	Non-Latino Korean only, 18 to 64 years
			24	Non-Latino Korean only, 65 years or older
			31	Non-Latino Filipino only, under 12 years
			32	Non-Latino Filipino only, 12 to 17 years
			33	Non-Latino Filipino only, 18 to 64 years
			34	Non-Latino Filipino only, 65 years or older
			41	Non-Latino Vietnamese only, under 12 years
			42	Non-Latino Vietnamese only, 12 to 17 years
			43	Non-Latino Vietnamese only, 18 to 64 years
			44	Non-Latino Vietnamese only, 65 years or older
			51	Other or non-Asian only, under 12 years
			52	Other or non-Asian only, 12 to 17 years
			53	Other or non-Asian only, 18 to 64 years
			54	Other or non-Asian only, 65 years or older

See note at end of table.

Table 7-1. Definitions of the dimensions used in raking (continued)

Dimension	Level	Description	Categories	
8	Stratum (collapsed where necessary)	Race/ethnicity (3) x Age groups (4)	11	Latino, under 12 years
			12	Latino, 12 to 17 years
			13	Latino, 18 to 64 years
			14	Latino, 65 years or older
			21	Non-Latino White, under 12 years
			22	Non-Latino White, 12 to 17 years
			23	Non-Latino White, 18 to 64 years
			24	Non-Latino White, 65 years or older
			31	Non-Latino Non-White, under 12 years
			32	Non-Latino Non-White, 12 to 17 years
			33	Non-Latino Non-White, 18 to 64 years
			34	Non-Latino Non-White, 65 years or older
9	Region (collapsed where necessary)	Education (4)	1	Not applicable (age < 18 years)
			2	Less than High School
			3	High School grad or GED recipient
			4	At least some college
10	Region (collapsed where necessary)	# Adults in HH (3)	1	0 or 1 adult,
			2	2 adults,
			3	3 or more adults
11	Region(collapsed where necessary)	Nonlandline telephone dimension		See Table 7-3

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Before raking the set of weights for the landline/list and combined landline/list/cell phone samples, dimensions with levels or cells with fewer than 50 respondents were collapsed with "adjacent" cells. In dimensions 1, 2, 5, and 8 the collapsed cells were created within the geographic regions shown in Table 7-2. As Dimension 5 was defined at the region level, some cells were collapsed across regions if the regions did not contain enough respondents. Dimensions 3, 6, and 7 were defined at the state level because there were too few respondents in many of the cells at lower geographic levels. Dimensions 9, 10 and 11 were defined at the region level because the control totals needed to create these cells (education and type of household defined by number of adults in the household) were not available at the county level. When collapsing the cells, we ensured that there was at least one cell or a group of cells within each self-reported stratum. In this way, the raked weights summed to the total number of persons in each stratum.

Dimensions 9, 10 and 11 were defined at the region level because the control totals needed to create these cells (education and type of household defined by number of adults in the household) were not available at the county level. When collapsing the cells, we ensured that there

was at least one cell or a group of cells within each self-reported region. In this way, the raked weights summed to the total number of persons in each stratum.

Table 7-2. Regions in California

Region	Sampling strata
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
Greater Bay Area	Santa Clara, Alameda, Contra Costa, San Francisco, San Mateo, Sonoma, Solano, Marin, Napa
Sacramento Area	Sacramento, Placer, Yolo, El Dorado
San Joaquin Valley	Fresno, Kern, San Joaquin, Stanislaus, Tulare, Merced, Kings, Madera
Central Coast	Ventura, Santa Barbara, Santa Cruz, San Luis Obispo, Monterey, San Benito
Los Angeles	Los Angeles
Other Southern California	San Diego, Orange, San Bernardino, Riverside, Imperial

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

For the area sample weights, control totals were created by restricting the files with the population totals to include adults in Los Angeles County or region 6. Dimension 1 was not used because it was redundant with dimension 3 that contains more detailed levels. Because the area sample size is smaller than the landline/list samples, more extensive collapsing was required.

7.3 Nonlandline Telephone Adjustments

CHIS 2007 included a cell phone sample to address the problem of traditional telephone samples where households without a landline telephone —including households with only cellular telephones— do not have a chance of being selected. The potential for bias from this undercoverage is related to the percentage of households without landline telephones and the differences in characteristics of the persons residing in households with landlines and those without.

CHIS 2005 was the first cycle to include a nonlandline adjustment focused on reducing the potential bias introduced by exclusion of wireless-only households from the survey. Like the nontelephone adjustment in CHIS 2001 and CHIS 2003, this adjustment was included as an additional raking dimension at the person level (dimension 11). The adjustment could not be done at the household level because it required data from the adult extended interview.

To achieve consistency across CHIS cycles, the combined landline/list sample weights include the same non-landline raking dimension implemented in 2005. This “no-landline” raking dimension was also used to create the combined landline/list/cell phone sample weights and area sample weights even though these weights included sampled persons from cell-only households. The main reason for the inclusion in raking is that this dimension does not control the total number of adults in cell-only households but the total of adults with characteristics related to this population. A better consistency among the set of weights is also achieved, and potential bias between estimates produced using the landline/list weights, combined landline/list/cell phone sample weights, and area weights can easily be identified because the weights have the same adjustments.

The goal of the nonlandline adjustment is to adjust the weights of adults in a weighting cell with a similar propensity of having a landline telephone. The variables used to create the raking cells were those identified by Blumberg et al. (2006) as good predictors of whether a household has a landline telephone. The control totals were derived for the same cells using the 2007 California Department of Finance (DOF) Population Estimates and the 2006 ACS public use micro data file (ACS-PUMS). Table 7-3 shows the definition of the 16 cells of dimension 11 used for the nontelephone adjustment in CHIS 2007. In CHIS 2007, this raking dimension was not created at the state level but included separate cells for Los Angeles County and the rest of California.

Table 7-3. Nonlandline telephone adjustment cell definition for CHIS 2007

Dimension 11 levels	Stratum	Household tenure	Age in years	Educational attainment	Number of adults in the household
1	Los Angeles	Own	0 to 17	NA	0 or 1
2		Rent	0 to 17	NA	0 or 1
3		Own	0 to 17	NA	2 or more
4		Rent	0 to 17	NA	2 or more
5		Own	18 to 30	Up to high school	NA
6		Own	31 to 64	Up to high school	NA
7		Own	65 and older	Up to high school	NA
8		Own	18 to 30	Greater than high school	NA
9		Own	31 to 64	Greater than high school	NA
10		Own	65 and older	Greater than high school	NA
11		Rent	18 to 34	Up to high school	NA
12		Rent	35 and older	Up to high school	0 or 1
13		Rent	35 and older	Up to high school	2 or more
14		Rent	18 to 34	Greater than high school	NA
15		Rent	35 and older	Greater than high school	0 or 1
16		Rent	35 and older	Greater than high school	2 or more
101	Remainder of California	Own	0 to 17	NA	0 or 1
102		Rent	0 to 17	NA	0 or 1
103		Own	0 to 17	NA	2 or more
104		Rent	0 to 17	NA	2 or more
105		Own	18 to 30	Up to high school	NA
106		Own	31 to 64	Up to high school	NA
107		Own	65 and older	Up to high school	NA
108		Own	18 to 30	Greater than high school	NA
109		Own	31 to 64	Greater than high school	NA
110		Own	65 and older	Greater than high school	NA
111		Rent	18 to 34	Up to high school	NA
112		Rent	35 and older	Up to high school	0 or 1
113		Rent	35 and older	Up to high school	2 or more
114		Rent	18 to 34	Greater than high school	NA
115		Rent	35 and older	Greater than high school	0 or 1
116		Rent	35 and older	Greater than high school	2 or more

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

7.4 Raking Factors

Table 7-4 shows the overall and relative raking adjustment factors for the adult, child, and adolescent weights for the landline/list sample. The overall adjustment factors were computed as the ratio of the control total to the sum of weights before raking. The factors in the table are larger than in previous cycles of CHIS. Since in 2007 the household weights were not poststratified to control totals as in previous cycles of CHIS, person-level raking also is adjusting for household undercoverage. It also adjusts for other sources of error such as reporting error and residual

nonresponse error. Thus, the raking factors cannot be used as a measure of person-level undercoverage. Nevertheless, they may be used as an indicator of which groups were harder to reach, or were less likely to complete the interview. Larger adjustment factors suggest relative undercoverage and smaller factors relative overcoverage.

Table 7-4. Overall adjustment raking factors for adult, child, and adolescent interviews by sample characteristics

Characteristic	Adult	Child	Adolescent
Total	1.453	1.549	1.480
Sex			
Male	1.563	1.547	1.503
Female	1.361	1.550	1.457
Age group			
Under 5 years		1.608	
6 – 11 years		1.493	
12 – 17 years			1.480
18-24 years	2.064		
25-29 years	2.440		
30-39 years	1.923		
40-49 years	1.516		
50-64 years	1.163		
65 years and over	0.957		
Race/Ethnicity			
Latino	1.995	1.786	1.825
Non-Latino			
White alone	1.187	1.211	1.109
African American alone	1.702	1.946	2.501
American Indian/Alaska Native alone	1.326	1.696	1.423
Asian alone	1.736	1.960	2.100
Native Hawaiian and Other Pacific Islander alone	2.398	2.068	1.807
Two or more races	0.913	1.173	0.691
Non-Latino Asian ethnic groups			
Chinese only	1.423	1.495	1.427
Korean only	1.583	1.738	2.325
Filipino only	2.548	2.671	3.610
Vietnamese only	2.334	3.522	4.911
Educational Attainment			
Not applicable (age < 18 years)		1.549	1.480
Less than High School,	2.158		
High School grad or GED recipient,	1.725		
Some college	1.279		
College degree or above	1.213		

Table 7-4. Overall adjustment raking factors for adult, child, and adolescent interviews by sample characteristics (continued)

Characteristic	Adult	Child	Adolescent
Household Tenure [a]			
Owner	1.328	1.393	1.319
Renter	1.718	1.780	1.842
Number of adults in the household [a]			
One	1.608	1.758	1.600
Two	1.362	1.485	1.421
Three or more	1.997	1.845	2.117
Number of children in the household [a]			
None	1.386		1.420
One	1.695	1.558	1.552
Two or more	1.798	1.545	1.621
Number of adolescents in the household [b]			
None	1.447	1.547	
One	1.507	1.556	1.524
Two or more	1.553	1.574	1.437

[a] Person level estimate by type of household

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 7-4 shows that the undercoverage for children is only slightly greater than the undercoverage for adults and adolescents. For adults, the adjustment factor for males is slightly larger than for females, which is common in household surveys. The factors also suggest undercoverage of younger adults.

One large adjustment factor is for persons who self-reported as having less than a high school education (2.158). The factors for the non-Latino Asian, non-Latino African American and non-Latino Native Hawaiian and Other Pacific Islander groups are also all larger suggesting potential undercoverage. Other factors worth noting include those for persons who self-reported as being two or more races. The small factors for these race groups suggests the CHIS 2007 estimates of persons of two or more races before raking are much higher than the corresponding DOF 2007 totals. It is likely that the way the race question is asked in CHIS (prompting for “any other race?”) encourages reporting of multiple races as compared to the Census methodology

7.5 Sources Used to Produce the Control Totals for CHIS 2007

In all cycles of CHIS considerable thought was given to the choice of data for the primary source of the control totals. In CHIS 2001, Census 2000 data were originally used because those data were recently compiled. However, as the cycles of CHIS move away from 2000, census

data do not reflect the current population as well. During the CHIS 2003 cycle, several sources for control totals were examined, including Census 2000 files (U.S. Census Bureau, 2001), the 2002 American Community Survey (ACS) (U.S. Census Bureau, 2003), and the 2003 California DOF Population Projections (State of California, Department of Finance, 2004). The DOF projections were settled on as the primary source for control totals in 2003 with the hope that they could also be used for future cycles of CHIS. The DOF provides Population Projections at the county level by race, ethnicity, gender and single age for each year from 2000 to 2050. The DOF also provides Population Estimates (State of California, Department of Finance, 2006a). The estimates are updated projections based on current birth and death data. The difference between the DOF projections and estimates is that the former are produced prior to the projected year and the latter after the estimated year. Therefore, the distributions of the DOF Population Estimates are more representative of the population.

Based on discussions with UCLA, the 2007 California DOF Population Projections poststratified to 2007 DOF Population Estimates were used as the primary source of control totals for the demographic control totals (i.e., raking dimensions defined by gender, race, ethnicity, age, and stratum) for CHIS 2007.

As in previous cycles of CHIS, the population totals had to be adjusted to remove the population living in group quarters who are not included in the survey, and household characteristics needed to be estimated. The 2000 Census files were used to compute the proportion of persons living in group quarters. The 2006 ACS files were also used as a source for educational attainment, household tenure, and household composition.

7.5.1 California Department of Finance Population Predictions

The DOF population projections are provided at the county level by gender, race/ethnicity and single age for each year as indicated in Table 7-5. The DOF population projections used the 2000 Census counts not adjusted for the Census 2000 undercount as the baseline. A baseline cohort-component method was used to project population estimates based on fertility/mortality rates and life expectancy by different race-ethnic groups and age cohorts. Special populations (prisons, colleges, and military installations) that have very different demographic and behavioral characteristics from the rest of the population were removed from the baseline and projected separately. However, the DOF held most of the special populations only at the year 2000

level. This factor played an important role in the assumptions made when removing the population living in group quarters from the control totals in CHIS 2007.

Table 7-5. Definition of counts available in the 2007 California DOF population projections files*

Variable	Available counts
Age groups (101)	Age 0 Age 1 . . . Age 100 or more
Sex (2)	Male Female
Race-ethnicity (6)	Latino, any race Non-Latino White alone Non-Latino African American alone Non-Latino American Indian/Alaska Native alone Non-Latino Asian alone Non-Latino Native Hawaiian and Other Pacific Islander alone Non-Latino Two or more races

* Available at the county level

Source: State of California, Department of Finance.

In previous years, the DOF provided an additional file with population counts for Latinos by race. This file was used to create the population counts by the full cross-tabulation age × sex × ethnicity × race. However, beginning in 2006, the DOF stopped providing this file. We used the corresponding file for 2005 to compute the proportion of race groups within Latinos and apply them to the total Latino population in the 2007 DOF file.

The main disadvantage of the DOF projections is the race categorization. The DOF population estimates follow the U.S. Office of Management and Budget (OMB) race definition known as "modified" race with no separate population counts for "other" race. The DOF estimates comply with the OMB 1997 revised standards for collection, tabulation, and presentation of federal data on race and ethnicity (Office of Management and Budget, 1997). The revised OMB standards identify only five main racial categories (White, Black or African American, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander) and combinations of these categories. In CHIS, respondents who could not identify themselves as any of the five OMB race categories could answer with a sixth category, "some other race," consistent with the 2000 Census data collection method. Recoding of "other race" for CHIS 2007 largely followed Census procedures (see *CHIS 2007 Methodology Series: Report 3 - Data Preparation*).

In order to use the DOF estimates, any sampled person who reported "other race" (alone or in combination with another race) had to be recoded into the OMB categories. This would have required the imputation of an OMB race category for 6,430 persons (9.95 percent of the CHIS 2007 sample) who self-reported "other race" only. As an alternative, a variable combining ethnicity with the OMB race that reduced the number of imputations was proposed and approved by UCLA. The recoding includes an additional level that arranges Latinos of any race into one group as shown in Table 7-6. Because most respondents who reported "other race" only were Latinos, the number of imputed records was reduced significantly to 127 persons (0.18 percent) who reported a non-Latino other race. The advantage of this additional variable is that it matches the categories of the population projections available in the DOF files. See Section 8.4 for additional details for the creation and imputation of this variable, OMBSRREO.

Table 7-6. Description of the variable using the OMB race definition

OMBSRREO	Definition
1	Latino
2	Non-Latino White alone
3	Non-Latino African American alone
4	Non-Latino Asian alone
5	Non-Latino American Indian/ Alaska Native alone
6	Non-Latino Pacific Islander alone
7	Non-Latino two or more races

The DOF population estimates include the population living in group quarters. Since the target population in CHIS 2007 excludes persons in group quarters, these persons were removed from the DOF population projections. At the time control totals were being developed, the DOF did not have separate projections for the population living in group quarters. The Census 2000 files were used to estimate the proportion of persons in group quarters, and these proportions were applied to the DOF estimates.

7.5.2 Census 2000 Files

The Census files were used to derive the control totals for the dimension defined by SPAs in Los Angeles and Health Regions in San Diego County (dimension 4 in Table 7-1). The proportions of the total population in those areas were computed from the 2000 Census files. This assumes that the proportion in these areas with respect to the county did not change between 2000

and 2007[10]. The Los Angeles SPAs and San Diego Health Regions were both defined in terms of Census Tracts.

7.5.3 American Community Survey for California

The American Community Survey (ACS) is a nationwide survey that provides current and detailed demographic, social, economic, and housing data. It is a critical element in the Census Bureau's reengineered 2010 Census plan as it will replace the decennial census long form in future censuses. The 2006 ACS produces population and household estimates for a limited number of characteristics at the state level and for over 800 geographical areas excluding the group quarters populations. There is a 2006 ACS public use micro data file (ACS-PUMS) for California that provides household and population estimates at the state level. For selected counties and large communities, custom tables for a subset of estimates can be downloaded from the Census website (http://www.census.gov/acs/www/). The 2006 ACS includes population estimates for 24 CHIS 2007 strata, but not for the SPAs in Los Angeles County, San Diego County Health Regions, or many small counties.

Although the 2004 ACS was not used as the primary source of population control totals in CHIS 2007, it was used to compute proportions by educational attainment and type of household (tenure and number of adults in the household) at the region level as these variables were not available in the DOF files. These proportions were applied to the 2007 DOF total population counts to derive the control totals for the raking dimensions defined by these characteristics (dimensions 9, 10, and 11 in Table 7-1). The proportions were calculated at the region level allowing use of the ACS data. Applying the 2007 factors assumed that there were no changes in the population proportions between 2006 and 2007 for these variables.

7.6 Producing the Control Totals for CHIS 2007

As with previous cycles of CHIS, the derivation of the control totals was a challenging task in 2007. It involved the selection of the sources of control totals, determining the number of dimensions, and computing the control totals. It also had an impact on the set of variables to be imputed. In CHIS 2007, there were 11 raking dimensions. Deriving the control totals for each

[10]The population in group quarters was removed from these areas and the county before computing the proportions.

dimension independently could lead to inconsistencies between totals across the dimensions and this would cause problems in the raking process.

To overcome these difficulties, we used a procedure developed for CHIS 2003 in which the control totals for most of the dimensions were computed simultaneously. In this approach, a file was created with totals for all the possible combinations of the levels from most of the raking dimensions in the source files. These totals were then adjusted to remove the population living in group quarters. In the final step, the file was summarized by aggregating the totals by raking dimension. Because all totals were produced from the same file, there were no inconsistencies (the sum across dimensions was constant and the relationship between dimensions using the same variables such as age was fixed) among the dimensions. The details of this procedure are described in the following sections.

As the first step when computing control totals, the population living in group quarters was removed from the population counts. This is a straightforward process when counts of persons in group quarters for all variables and geographic levels are available. However, this information was not available in the DOF files. By assuming that the proportion of the population in group quarters did not change between 2000 and 2007, the Census 2000 SF1 file could be used to compute these proportions. This assumption is the same one used by the California DOF for its population projections.

In past cycles of CHIS, two problems occurred when computing the percentage of the population living in group quarters using the Census SF1 file. The first was the limited number of group quarter counts that can be produced from the SF1 file. Counts are available only by stratum (44) × age group 1 (3) × sex (2) × race (7) and by stratum (44) × age group 1 (3) × sex (2) × ethnicity (3) as defined in Table 7-7. Other counts included totals by stratum (44) and single year of age.[11] The file could not be used to produce population counts by single year of age by the cross-tabulation of race and ethnicity. The file with counts by single age was used to compute the population in group quarters for younger and older children and adolescents, e.g., and stratum (44) × age group 2 (5) where age group 2 is defined in Table 7-7. As in previous cycles the assumption was made that the distribution of the population in group quarters is uniform among the age

[11] Census 2000 totals for the population in group quarters by single age were not used in previous cycles. These counts were used for the first time in CHIS 2007 for more accurate control of the proportion in group quarters of two age groups for children (0 to 5 and 6 to 12 years old) and adolescents. The proportion of the population in group quarters is not uniform across age groups; for example, the proportion of adolescents in group quarters is much larger than the proportion of younger children.

groups; for example, if the percentage of persons 65 or older in group quarters is 1.56 percent, then 1.56 percent of persons 68 years old are assumed to be in group quarters.

The second problem was that the group quarter population counts from the SF1 file are defined for the seven race categories shown in Table 7-7 and not the six OMB race groups used in the DOF file (see Table 7-5). To address this problem, we assumed that the distribution of persons in group quarters by ethnicity (Latino or non-Latino) was also the same within race. For example, if 1.42 percent of the African American population is in group quarters, then 1.42 percent of both Latino African Americans and non-Latino African Americans are assumed to be in group quarters.

Table 7-7. Definition of levels of variables for group quarters populations in the Census 2000 SF1 file

Characteristics	Available counts
Stratum (44)	Counties or combinations of multiple counties defined in CHIS 2007
Age group1 (3)	Less than 18 years old
	18 to 64 years old
	65 years old or older
Age group2 (5)	Less than 6 years old
	6 to 11 years old
	12 to 17 years old
	18 to 64 years old
	65 years old or older
Sex (2)	Male
	Female
Race (7)	White alone
	African American alone
	American Indian/Alaska Native alone
	Asian alone
	Native Hawaiian and Other Pacific Islander alone
	Other race alone
	Two or more races
Ethnicity(3)	Latino
	Non-Latino White alone
	Other

Source: U.S. Census Bureau, Census 2000.

Using this assumption, the percentage of the population not living in group quarters in 2000 was computed as follows. A file with 2000 population totals, T_{rc}^{2000}, was created by summarizing the 2000 SF1 into 22,176 cells denoted *rc*, where *r* denotes race and *c* is the cross-tabulation of $\text{stratum}(44) \times \text{ethnicity}(2) \times \text{age}(18) \times \text{gender}(2)$. The 18 levels of age (see Table 7-8) corresponded to the cross-tabulation of the levels of age available in the DOF data files and in the definition of the raking dimensions. An advantage of summarizing the file by the levels of *c* was

the smaller size of the file (i.e., the file contains population totals by the age groups rather than single age). Note that any age group, race, or ethnicity as defined in the raking dimensions could be created by combining the *c* cells.

We defined the cells *rc* as the cross-tabulation of race and the cell *c* as follows:

$$rc = \text{race}_{\overline{OMB}}(7) \times c \,,$$

where the subscript $\overline{OMB}$ refers to the non-OMB race classification that includes a category for "some other race" available in the SF1 file as shown in Table 7-7.

Table 7-8. Age levels corresponding to the cross-tabulation of the DOF data files and the definition of the raking dimensions

Age group(*i*)	Description
1	0 to 3 years old
2	4 to 5
3	6 to 7
4	8 to 11
5	12 to 14
6	15 to 17
7	18 to 24
8	25
9	26 to 29
10	30
11	31 to 37
12	38 to 39
13	40 to 45
14	46 to 49
15	50 to 53
16	54 to 64
17	65 to 77
18	78 plus

7.6.1 Removing the Population Living in Group Quarters

We now review how the group quarter population was removed from the DOF files. Define $T_{rc}^{2000\,\overline{GQ}}$ as the 2000 population total that excludes the population in group quarters in cell *rc*. The totals $T_{rc}^{2000\,\overline{GQ}}$ were computed by raking the totals T_{rc}^{2000} to three control totals for the population not living in group quarters. Let $D1_{m}^{2000\,\overline{GQ}}$ be the control total for the fist raking dimension computed as

$$D1_m^{2000\,\overline{GQ}} = D1_m^{2000} - D1_m^{2000\,GQ},$$

where $D1_m^{2000}$ is the 2000 total population, $D1_m^{2000\,GQ}$ is the 2000 population total living in group quarters, and *m* is the raking cell defined as $m = \text{strata}(44) \times \text{race}_{\overline{OMB}}(7) \times \text{age group } 1(3) \times \text{sex}(2)$.

In the same way, let $D2_n^{2000\,\overline{GQ}}$ be the control total for the second raking dimension for cell *n*, where *n* is defined as $n = strata(44) \times ethnicity(3) \times age\ group\ 1(3) \times sex(2)$ as in the SF1. Let $D3_p^{2000\,\overline{GQ}}$ be the control total for the third raking dimension for cell *p*, where *p* is defined as $n = strata(44) \times age\ group\ 2(5)$ as in the SF1.

Note that $D1_m^{2000\,GQ}$, $D2_n^{2000\,GQ}$, and $D3_p^{2000\,GQ}$ are the 2000 population totals living in group quarters available in the SF1 file. By using raking we ensured that all totals, $T_{rc}^{2000\,\overline{GQ}}$, were consistent and they summed to the control totals.

After raking, the proportion of the 2000 population not living in group quarters in cell *rc* was computed as

$$p_{rc}^{2000\,\overline{GQ}} = \frac{T_{rc}^{2000\,\overline{GQ}}}{T_{rc}^{2000}}.$$

Assuming that the proportion of the population not living in group quarters did not change between 2000 and 2007 within cell *rc*, the proportion $p_{rc}^{2000\,\overline{GQ}}$ could be used to compute $T_{rc}^{2007\,\overline{GQ}}$ defined as the 2006 total population not living in group in cell *rc*, as

$$T_{rc}^{2007\,\overline{GQ}} = p_{rc}^{2000\,\overline{GQ}} * T_{rc}^{2007},$$

where T_{rc}^{2007} is the 2007 total population from the 2007 California DOF file in cell *rc*. However, T_{rc}^{2007} could not be computed using the DOF file due to differences in race categorization between the SF1 and the DOF projection. Instead, the 2007 population estimates, $T_{sc}^{2007\,OMB}$, were available in the DOF file for 19,008 cells (labeled *sc*) defined using the OMB race categories. The cells *sc* were defined by the cross-tabulation of $sc = \text{race}_{OMB}(6) \times c$, where the subscript *OMB* refers to the OMB race groups that exclude the "some other race" category as shown in Table 7-9, and *c* is defined as before.

Table 7-9. OMB race categories available in the California DOF files

$\text{race}_{\text{OMB}}(s)$	Description
1-W	OMB White alone
2-AA	OMB Black or African American alone
3-AI	OMB American Indian or Alaska Native alone
4-AS	OMB Asian alone
5-PI	OMB Pacific Islander Native Hawaiian alone
6-TM	OMB Two or more races

In order to examine the relationship between the totals T_{sc}^{OMB} and T_{rc}, consider the following summation:

$$T_c^{OMB} = \sum_s T_{sc}^{OMB} = T_{Wc}^{OMB} + T_{AAc}^{OMB} + T_{AIc}^{OMB} + T_{ASc}^{OMB} + T_{PIc}^{OMB} + T_{TMc}^{OMB} .$$

In the same way, the total population in a cell *c* can be represented by non-OMB race groups as

$$T_c = \sum_r T_{rc} = T_{Wc} + T_{AAc} + T_{AIc} + T_{ASc} + T_{PIc} + T_{Oc} + T_{TMc} .$$

The assignment of OMB race was done within cell *c*; in other words, the total population in the cell *c* stays constant. That is

$$T_c = \sum_r T_{ri} = T_c^{OMB} = \sum_s T_{sc}^{OMB} .$$

When assigning an OMB race value, persons who reported "some other race" alone were assigned one of the OMB race categories. Persons who reported two races, one being "other race," kept the OMB race category but dropped "other race." In other words, they were assigned a single OMB race. Persons who reported more than two races, one of these being "other race," were still considered as having multiple races (the "other race" removed).

In order to illustrate the reallocation, consider the Asian group (ignoring the stratum, age group, sex, and ethnicity components of the cell),

$$T_{ASc}^{OMB} = T_{ASc} + p_{AS\_O}^{OMB} * T_{Oc} + p_{AS\_TM}^{OMB} * T_{TMc} ,$$

where

T_{ASc}^{OMB} is the total number of Asians (OMB definition);

T_{ASc} is the total number of Asians (non-OMB definition);

T_{TMc} is the total number of persons with two or more races (non-OMB definition);

$p_{AS\_O}^{OMB}$ is the proportion of persons with some other race alone who were coded as Asian alone when assigning the OMB definition; and

$p_{AS\_TM}^{OMB}$ is the proportion of persons with two or more races who are coded as Asian alone when assigning the OMB definition.

In other words, the OMB Asian alone population (T_{AScc}^{OMB}) is composed of the original non-OMB Asian-alone total (T_{AScc}), the portion of the population who reported "some other race" alone that is allocated to OMB Asian ($p_{AS\_O}^{OMB} * T_{Oc}$), and the population who reported non-OMB Asian-alone and "some other race." Figure 7-1 visualizes how the OMB Asian-alone population is formed, where T_{AS}^{OMB}, T_{AS}, and T_{TM} are defined above and T_O is defined as the group who reported "other race" only (omitting the subscript *c* for convenience).

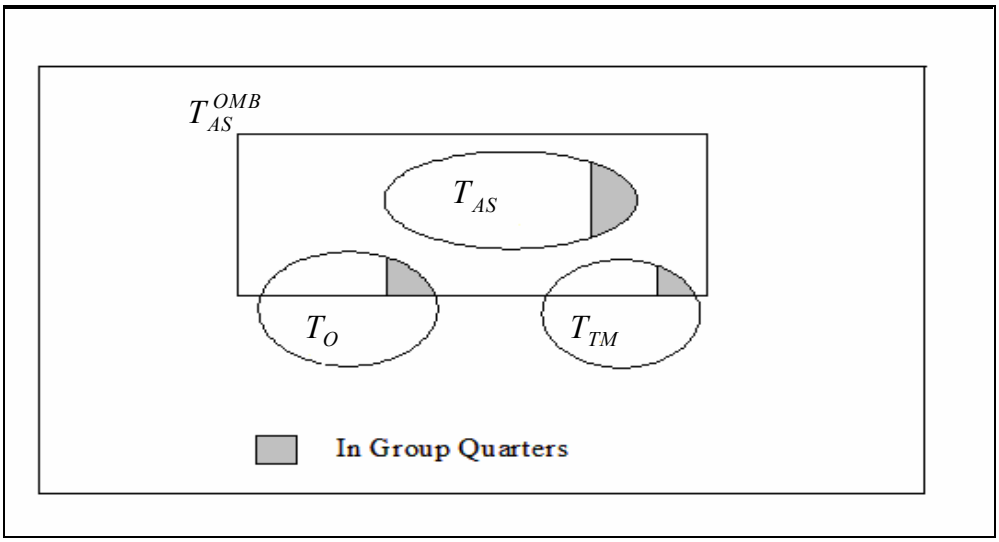


Figure 7-1. Relationship between OMB Asian alone and non-OMB groups

In this example, the proportion of the population in group quarters was known for the non-OMB Asian alone group. In order to compute the proportion of the population not in group quarters for OMB Asian alone we assumed the same proportion holds for the members that were being reclassified into the OMB race group. That is

$$\frac{T_{AS}^{OMB\,\overline{GQ}}}{T_{AS}^{OMB}} \approx \frac{T_{AS}^{\overline{GQ}}}{T_{AS}} = \frac{T_{AS}^{\overline{GQ}} + T_O^{\overline{GQ}} + T_{TM}^{\overline{GQ}}}{T_{AS} + T_O + T_{TM}},$$

only for $O \in AS$ and $TM \in AS$, (i.e., OMB race assignment to *AS*).

Generalizing these results to the other groups, the proportion of the population not in group quarters, $p_{rc}^{\overline{GQ}}$, can be computed as

$$p_{rc}^{\overline{GQ}} = \frac{T_{rc}^{\overline{GQ}}}{T_{rc}} \approx \frac{T_{sc}^{OMB\,\overline{GQ}}}{T_{sc}^{OMB}} = p_{sc}^{\overline{GQ}}$$

Under the assumption that the proportion of the population not living in group quarters did not change between 2000 and 2007 the proportion was computed as

$$p_{sc}^{2007\,\overline{GQ}} = p_{sc}^{2000\,\overline{GQ}} = \frac{T_{rc}^{2000\,\overline{GQ}}}{T_{rc}^{2000}}.$$

The proportion $p_{sc}^{2007\,\overline{GQ}}$ was used to compute the 2007 total population not living in group quarters in cell *sc,* $T_{sc}^{2003\,OMB\,\overline{GQ}}$, defined using the OMB race categories, as follows:

$$T_{sc}^{2007\,OMB\,\overline{GQ}} = p_{sc}^{2000\,\overline{GQ}} * T_{sc}^{2005\,OMB} = \frac{T_{rc}^{2000\,\overline{GQ}} * T_{sc}^{2005\,OMB}}{T_{rc}^{2000}}.$$

where T_{rc}^{2000} is computed using the SF1 file, $T_{sc}^{2007\,OMB}$ using the 2007 DOF file and $T_{rc}^{2000\,\overline{GQ}}$ is the 2000 population in cell *rc* not in group quarters, as defined earlier. The 2007 total population not living in group quarters in California is computed as

$$T^{2007\,\overline{GQ}} = \sum_s \sum_c T_{sc}^{2007\,OMB\,\overline{GQ}}.$$

Table 7-10 shows the total population in the 2007 DOF file and the estimated total (and percentage) of the population living in group quarters.

Table 7-10. Population in California in 2007 by group quarter status

Type	Population	%
In group quarters	819,754	2.42
Not in group quarters	33,051,894	97.58
Total	33,871,648	100.00

Source: California Department of Finance, 2007.

7.6.2 Computing the Control Totals

The totals $T_{sc}^{2007\,OMB\,\overline{GQ}}$ were summarized in order to compute the control totals for dimensions 1, 2, 3, 5, 6, and 8. For dimension 7, defined for Asian ethnic groups, the control totals

were derived using the same totals but for Asian only. Using the Census 2000 SF1 files, we computed the percentages for the Asian groups in Table 7-11. The percentages of the Asian groups by ethnicity (Latino, non-Latino) were computed using the 2006 ACS-PUMS file. It was assumed that there were no changes in the distribution of the Asian groups between 2006 and 2007. These percentages were applied to the 2007 DOF projections.

Table 7-11. Census 2000 SF1 Asian groups

Asian Group	Description
1	Chinese alone
2	Korean alone
3	Filipino alone
4	Vietnamese alone
5	Other Asian ethnic group alone

The creation of dimension 4, defined by SPAs in Los Angeles County and Health and Human Services Agency (HHSA) Service Regions in San Diego County, used information from the Census 2000 SF1. The Los Angeles County Department of Health (LACDH) produced a listing of Census tracts by SPA. The 2000 SF1 file was used to compute the percentages of the population in the SPAs by aggregating population counts in the Census tracts. This percentage was applied to the total 2007 DOF population total (excluding group quarters) to produce the controls for dimension 4. A similar procedure was used for San Diego County Health Regions.

For dimensions 9 (adult's education attainment), 10 (number of adults in the household), and 11 (nontelephone adjustment), the percentages of the population were computed using the 2006 ACS-PUMS and then applied to the 2007 DOF population total (excluding group quarters). The underlying assumption was that there were no changes in the distribution of the population between 2006 and 2007.

8. IMPUTATION PROCEDURES

In any household survey, both unit and item nonresponse are virtually unavoidable. We have described how weighting adjustments have been used to compensate for unit nonresponse in CHIS 2007. *CHIS 2007 Methodology Series: Report 4 – Response Rates* discusses unit nonresponse in detail. This chapter focuses on item nonresponse and the imputation for missing responses of the variables used in weighting. The imputed values were needed in the last stages of the weighting process, and only interviews that were considered completed units were subject to imputation. The percentage of missing data and consequent imputation for virtually all of these items is small.

Section 8.1 describes the imputed variables and reviews the different types of imputation techniques used to fill in the missing data. The two imputation techniques employed in CHIS 2007 are random allocation and hot-deck imputation. Sections 8.2 through 8.4 discuss the imputation process for all imputed variables separately. The last section lists the geographic location variables for CHIS 2007. UCLA derived these variables after geocoding the geographic information either collected during the interview (address of respondent, nearest street intersection, self-reported county) or attached to the sample telephone (address for numbers that were mailable or ZIP Code covered by the telephone exchange).

8.1 Imputed Variables and Methods

Table 8-1 lists the variables imputed for weighting in CHIS 2007. With a few exceptions, the same variables were imputed in previous cycles of CHIS. In 2007, two additional variables were imputed in order to create a variable that defines a household by type of telephone service available to its residents. As noted above, the level of missing data is relatively small. The specific percentages of missing data are given later in the chapter. When the amount of missing data is small and assuming that the data are missing at random (i.e., the missing data have the same distribution as those with complete data within groups defined for imputation), then the bias of the estimates due to the missing data should be relatively small. The imputations may also increase the variance of the estimates, but this effect should be negligible given the low rate of missing data. A flag indicating if the response is imputed accompanies every value.

Table 8-1. Description of imputed variables

Variable name	Description	Interview items	Variable type
SRAGE	Self-reported age	AA2, CA3, TA2, KAA2	Demographic
SRSEX	Self-reported sex	AA3, CA1, TA3, KAA3	Demographic
SRTENR	Self-reported household tenure	AK25, KAK25	Socio-economic
SREDUC	Self-reported educational attainment	AH47, KAK47	Socio-economic
SRH	Self-reported Latino	AA4, CH1, TI1	Ethnicity
SRW	Self-reported white	AA5A_6, CH3_6, TI2_6	Race
SRAA	Self-reported African American	AA5A_5, CH3_5, TI2_5	Race
SRAS	Self-reported Asian	AA5A_4, CH3_4, TI2_4	Race
SRAI	Self-reported American Indian/ Alaska Native	AA5A_3, CH3_3, TI2_3	Race
SRPI	Self-reported Native Hawaiian and Other Pacific Islander	AA5A_1, AA5A_2, CH3_1, CH3_2, TI2_1, TI2_2	Race
SRO	Self-reported Other race	AA5A_7, CH3_7, TI2_7	Race
OMBSRREO	OMB self-reported race/ethnicity		Race/ Ethnicity
OMBSRASO	OMB self-reported non-Latino Asian group	AA5E_1- AA5E_18, TI7_1- TI 7_18, CH7_1- CH7_18	Race/ Ethnicity
CELLPHONE	Cell/Wireless telephone service in household	SC9, SC5, AM	Telephone usage
LANDLINE	Landline telephone service in household	SC9	Telephone usage

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

In CHIS 2007 and previous cycles, random allocation and hot-deck imputation were used to fill in the missing responses. The first imputation technique is a random selection from the observed distribution. This method is used only when the item missing rate is very small.

The second technique, hot-deck imputation, was used to impute race and ethnicity (including the OMB race-ethnicity variables) as well as household tenure and educational attainment in the previous cycles of CHIS. The hot-deck approach is probably the most commonly used method for assigning values for missing responses in large-scale household surveys (Sande, 1983; Ford, 1983). 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. In order to carry out hot-deck imputation for CHIS 2007, the respondents to an item form a pool of donors while the nonrespondents are a group of recipients. A recipient is matched to the subset pool of donors with the same characteristics. The recipient is then assigned a randomly imputed value from one of the donors in the pool. Once a donor is used, it is removed from the donor pool.

8.2 Self-Reported Sex and Age

The percentage of cases where either sex or age was missing in CHIS 2007 is very small across all samples (landline, surname list, cell phone, and area) and types of extended interviews (adult, child, and adolescent). Table 8-2 summarizes the number of cases that were imputed for sex and age. The sex of only five children and one adult was imputed and no adolescents were missing self-reported sex. For children, the missing data for sex was imputed randomly. A random number was generated for the five missing values. Two children were imputed as male and three female. The sex of the adult was imputed using roster information obtained during the screener interview.

Age was imputed in 146 cases in CHIS 2007 across all samples. A hierarchical process was followed to impute the 130 missing self-reported age values for adults in the landline and list samples. The process used the values for self-reported age (question AA2 on the adult interview), the self-reported adult age range (question AA2A on the adult interview) asked when the adult refused to provide a specific age, the proxy-reported adult age collected during the child-first interview (question KAA2) if available, and the adult age collected during the screener interview (question ADULTAGE on the screener interview). The single respondent in the area sample that did not provide a specific age did provide an age range and had their age randomly imputed within that range.

Table 8-2. Number and percentage of completed interviews with missing self-reported sex and age by sample type

Sample Person type	Number completed	Number missing sex	% missing sex	Number missing age	% missing age
Landline/List					
Adult	49,242	1	0.00	130	0.26
Child	9,913	5	0.05	15	0.15
Adolescent	3,638	0	0.00	0	0.00
Total	62,793	6	0.01	145	0.23
Area	0				
Adult	981	0	0.00	1	0.10
Cell Phone	0				
Adult	825	0	0.00	0	0.00
Overall Total	64,599	6	0.01	146	0.23

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

The missing age for adults was imputed as follows. First, if an adult had a missing self-reported age, we checked whether the adult age was collected from a proxy adult in the child-

first interview. If age was reported, this age was assigned to the sampled adult. If an age was not reported in the child-first interview, the screener age for the sampled adult was checked. If the screener age was within an age range given by the sampled adult, then screener age was used as the imputed age. If the age range was not reported, then the screener age was used. However, if the screener age was outside the reported age range, then age was randomly selected using the distribution of self-reported age within the reported age range. If no age was collected during the screener but an age range was reported, then age was randomly imputed from the distribution of self-reported age within the reported age range. If no information on age was available from any source, then age was randomly imputed using the distribution of self-reported age of all adult respondents. As an example, assume an adult respondent did not report an age but reported an age range of 40 to 44. Assume also that the proxy reported adult age in the child-first interview was 38 and the age collected in the screener interview when the adult was sampled was 38. This situation could result if the proxy misreported the sampled adult age in both the screening interview and the child-first interview. For this case, the adult age would be imputed using the distribution of the self-reported age of adults age 40 to 44. Assume that the distribution of adult age is such that 41 percent of sampled adults were age 41 or less and 62 percent were age 42 or less. If the random number assigned to the adult had a value of 0.44 then the adult's imputed age would be 42 years old.

8.3 Household Tenure and Educational Attainment

Household tenure and the adult respondent's educational attainment were used to create raking dimensions 9 and 11. Household tenure had 340 missing responses (0.90 percent), and educational attainment had 182 missing (0.30 percent).

Hot-deck imputation was used to impute missing values for these two variables. The search algorithm CHAID (Kass, 1980) was used to create the hot-deck cells using the variables available for both donors and recipients found to be good predictors. A donor was then randomly drawn from the cell and its value for the variable being imputed was assigned to the recipient. Table 8-3 shows the variables considered in CHAID to create the hot-deck cells for educational attainment and household tenure. Table 8-4 shows the distribution of the imputed cases by sample type. When calculating the percentages, the denominator for educational attainment is the number of adults in a given education category, and for tenure the denominator is all adults who own or rent.

Table 8-3. Variables used to define hot-deck cells for the imputation of education attainment and household tenure

Variable Name	Description
Educational Attainment	
SRSEX	Self-reported sex
SRRACE_O	Self-reported race
SRH	Self-reported ethnicity
SRAGE	Self-reported age
ADLTFLG	Number of adults in the household
CHLDFLG	Children present in the household
TEENFLG	Adolescents present in the household
POVERTY	Poverty
P_GRAD	Percent college graduates in exchange
P_OWN	Percent home owners in the exchange
P_BLACK	Percent African Americans in the exchange
P_HISP	Percent Latinos in the exchange
CREGION	California Regions
Household Tenure	
ADLTFLG	Number of adults in the household
CHLDFLG	Children present in the household
TEENFLG	Teens present in the household
P_GRAD	Percent college graduates in exchange
P_BLACK	Percent African Americans in the exchange
P_HISP	Percent Latinos in the exchange
P_OWN	Percent home owners in the exchange
POVERTY	Poverty
CREGION	California Regions

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 8-4. Counts and percentages of imputed self-reported education attainment and household tenure

	Adult Interviews					
	Sample type					
	Landline/list		Area		Cell phone	
	Count	Percentage	Count	Percentage	Count	Percentage
Self-reported Education Attainment						
Under 18 years of age	NA		NA		NA	
Less than HS, 18 years of age or older	30	0.64	2	1.02	1	1.43
High School (or equivalent), 18 years of age or older	47	0.43	1	0.39	1	0.39
Some college, 18 years of age or older	45	0.32	1	0.41	2	0.75
BS and above, 18 years of age or older	54	0.27	0	0.00	0	0.00
Total	176	0.36	4	0.41	4	0.48
Self-reported Household Tenure						
Owner	175	0.51	4	0.97	0	0.00
Renter	156	1.04	1	0.18	4	0.68
Total	331	0.67	5	0.51	4	0.48

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

8.4 Self-Reported Race and Ethnicity

As described in Chapter 7, the person weights were raked to control totals from the 2007 California DOF Population Estimates. The California DOF complies with the OMB 1997 revised standards for collection, tabulation, and presentation of federal data on race and ethnicity. The revised OMB standards identify only five main racial categories and combinations of these categories. The main categories are White, Black or African American, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander. Census 2000 allowed a sixth race category ("some other race") for respondents who could not identify with any of the five OMB race categories. Because all public release files of the Census 2000 include six race categories, the Census Bureau released a special file called Modified Race Data Summary file (MRDSF) with 2000 population counts by the five OMB race categories (U.S. Census Bureau, 2002). To produce this file the Census Bureau implemented special procedures to assign and impute an OMB race to those who reported "some other race." The California DOF Estimates used the 2000 MRDSF as the baseline for the time series; as a result, the DOF Estimates include only counts by the five OMB racial categories by county.

Following a procedure similar to the Census 2000, respondents who could not identify themselves as any of the five OMB race categories could answer "some other race" in CHIS. In order to use the DOF estimates as control totals, any sampled person who reported "some other race" (alone or in combination) had to be recoded into one or more of the OMB categories. OMB race was missing 6,430 persons (9.95 percent) in CHIS 2007. After examining the procedures used by the Census to assign an OMB race, we determined that the assignment of OMB race could not be implemented using the available variables in CHIS 2007 as in Census 2000, because the number of CHIS cases in the geographic area (i.e., stratum) by Latino origin[12] cells is not large enough to guarantee a good assignment. The same situation occurred in previous cycles. To reduce the number of records to be imputed, a combined race/ethnic variable (OMBSRREO) that assigned Latinos regardless of race into one group was proposed and approved. The levels of the variable OMBSRREO are given in Table 8-5.

[12] Donors and donees must match on the specific Latino origin (Not Hispanic; Mexican; Puerto Rican, Cuban, Central American and Dominican; South American; Other Spanish).

Table 8-5. OBM race/ethnicity groups (OMBSRREO)

OMBSRREO	Description
1	Latino
2	Non-Latino White
3	Non-Latino African American
4	Non-Latino American Indian Alaskan Native
5	Non-Latino Asian
6	Non-Latino Pacific Islander Native Hawaiian
7	Non-Latino two or more races

By creating a separate group for Latinos, a valid value of OMBSRREO was missing for only 127 persons (0.18 percent) who self-reported as non-Latino and "some other race" alone[13] in 2007. The reduction in the number of cases is because most of the people who report other race were Latino. Using a variable that combined race-ethnic groups with one level of OMBSRREO for Latino eliminated the need to impute for 6,303 cases who reported Latino "other race" alone.

For continuity with the race and ethnicity variables created since 2001 (see Table 8-1), the same variables were created and imputed in 2007. We refer to these variables as the "regular" single race and ethnicity variables. The OMB race-ethnicity variable OMBSRREO was created using these regular race and ethnicity variables after imputation. Section 8.4.1 describes the imputation of the regular race and ethnicity variables while Section 8.4.2 describes the creation and imputation of the OMB race variable. Section 8.4.3 discusses the creation and imputation of self-reported Asian ethnic groups.

8.4.1 Imputation of Single Self-Reported Race and Ethnicity

While the procedures used to impute for missing values of sex and age were relatively straightforward, self-reported race and ethnicity presented a greater challenge. Different imputation methods were considered before choosing the final approach. One approach that was considered, but not adopted, was to use the self-reported race and ethnicity of a respondent to impute for any other sampled person with missing values for these items within the household. The reason this approach was not used in any cycle of CHIS is the realization that the method does not account for households with persons of more than one race and ethnicity.

Instead a hot-deck imputation procedure was developed to deal with the diversity of race and ethnicity within households. Before describing the hot-deck approach, some special

[13] This includes records imputed as non-Latino "other" from the regular CHIS 2005 race imputation.

features of the race and ethnicity items are worth noting. First, although race is a series of items with subparts, the items we deal with are only those that classify a person as White, African American, Asian, American Indian/Alaska Native, Pacific Islander, or other. Also, these items are treated as either all reported or all missing. In very few cases there were missing values for one of the races but not others, but the data preparation staff was able to replace these missing values using interviewer comments. Finally, some missing values were assigned deterministically based on other items such as country of origin. These deterministic imputations were flagged like all other imputations.

Table 8-6 shows the number and percentage of cases with imputed values by type of extended interview (adult, child, and adolescent). The first columns are those cases where race is imputed, and the next set of columns is for cases where ethnicity is imputed.

Table 8-6. Number and percentage of imputed interviews with missing self-reported race and/or ethnicity

Sample type	Imputed race*		Imputed ethnicity	
Type of interview	Count	%	Count	%
Landline/list				
Adult	1,192	2.42	146	0.39
Child	482	4.86	44	0.44
Adolescent	235	6.46	27	0.74
Area				
Adult	46	4.69	1	0.10
Cell phone				
Adult	42	5.09	2	0.24
Total	1,626	2.78	195	0.34

* At least one value of race was imputed.
Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

The hot-deck imputations were done separately by the completed extended interview structure of the household. In general, the imputation procedure was done at the household level and handled households with the fewest missing values first and then moved to the cases with more missing values. The simplest household structure is where only an adult was interviewed (versus a household with an adult and an adolescent and/or a child). A household with only one adult with missing ethnicity was imputed before a household with only an adult that had both missing race and ethnicity.

The patterns of missing data for race and ethnicity varied by the structure of the household. For the simple case where only an adult was interviewed, the donors were selected from other adult-only households. If the adult was missing both race and ethnicity, both values were

imputed from the same donor. If the adult had a reported race but was missing ethnicity, then a donor with the same race (all six race values were placed into a vector and only adults with the exact same values could be donors) was randomly selected. For an adult with reported ethnicity and missing race, the same procedure was used; only adults in adult-only households with the same value of ethnicity could be donors. Whenever possible, the donors and the recipients were from the same sampling stratum. For cases where the pool formed in this way had too few donors, sampling strata were combined based on geographic and urban status. Once a donor was used, it was removed from the pool for all future hot deck runs.

The same principles were used for more complex household structures. In these cases, some households had missing race and ethnicity for all sampled persons, while in others one or more of the sampled persons might have a reported race and ethnicity. Various combinations, such as a reported ethnicity but not race, were also encountered. Separate hot deck runs were made to accommodate all of these situations. As an illustration, consider households where an adult and a child are interviewed. Assume the adult reported non-Latino ethnicity and Asian race and the child only reported non-Latino ethnicity but no race. The pool of donors for imputing the child's race consists of households where only an adult and a child were interviewed and where the adult reported non-Latino ethnicity and Asian race and the child reported non-Latino ethnicity. The households with other combinations of persons with missing race and/or ethnicity were imputed in an similar way. Table 8-7 shows the counts and percentages of imputed values by self-reported race and ethnicity and type of extended interview (adult, child, and adolescent).

Table 8-7. Counts and percentages of imputed interviews with missing self-reported race and ethnicity by type of extended interview

	Total		Extended interview type					
			Adult		Child		Adolescent	
	Count	%	Count	%	Count	%	Count	%
Self-reported race								
White alone	931	1.44	603	1.18	243	2.45	85	2.34
African American alone	59	0.09	30	0.06	18	0.18	11	0.30
Asian alone	58	0.09	31	0.06	19	0.19	8	0.22
American Indian/ Alaska Native alone	156	0.24	92	0.18	25	0.25	39	1.07
Pacific Islander alone	13	0.02	8	0.02	4	0.04	1	0.03
Other race alone	851	1.32	550	1.08	196	1.98	105	2.89
Two or more races	69	0.11	33	0.06	23	0.23	13	0.36
Total	2,124	3.29	1,347	2.64	528	5.33	262	7.20
Self Reported Ethnicity								
Latino	49	0.08	22	0.04	17	0.17	10	0.27
Non-Latino	171	0.26	127	0.25	27	0.27	17	0.48
Total	220	0.34	149	0.29	44	0.44	27	0.74
Completed interviews	64,599	100.00	51,048	100.00	9,913	100.00	3,638	100.00

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

8.4.2 Imputation of the OMB Race-Ethnicity Variable

The DOF control totals are defined in terms of OMB race categories for raking dimensions 5, 6, and 8. Persons who reported themselves as Latino "some other race" were assigned an OMB race following procedures similar to those used in Census 2000. Since the OMB assignment is done using the imputed regular single race variables, all sampled persons have nonmissing races values for variables SRW, SRAA, SRAI, SRAS, SRPI, and SRO.

The OMB race-ethnicity variable, OMBSRREO, was assigned as follows:

- If the person self-reported as Latino (SRH=1), the variable OMBSRREO was set to 1. This assignment is independent of the values of the race variables.
- If the person self-reported as non-Latino (SRH=2) and reported OMB race alone or in combination with one or more OMB races (e.g., White alone, White and Black or African American, White and Black or African American and American Indian and Alaska Native) then OMBSRREO was given the value 2, 3, 4, 5, 6 or 7 (see Table 8-5) depending on the values of SRW, SRAA, SRAI, SRAS, and SRPI. In other words, there is no modification of race for non-Latinos who reported a valid OMB race(s).

- If the person self-reported as non-Latino (SRH=2) and reported both an OMB race and "some other race" (SRO=1), then OMBSRREO was assigned using only the specified OMB race(s). For example, non-Latino White and some other race became non-Latino White alone. This scenario is an example of the differences between OMBSRREO and the regular race-ethnicity variables (SRH, SRW, SRAA, SRAI, SRAS, and SRPI). Persons who reported two races, with one of them "some other race" are considered as single race respondents based on the OMB definition.

After the race/ethnicity assignments were made, 127 persons (0.18 percent) remained with missing values of OMBSRREO. These persons self-reported as non-Latino and other race only (SRH=2 and SRO=1). The missing values were imputed using the same procedures used to impute the regular single race variables as described above. In this case, temporary OMB race variables named SRW2, SRAA2, SRAI2, SRAS2, and SRPI2 were created using the values of already imputed SRW, SRAA, SRAI, SRAS, and SRPI. The values of the temporary OMB race variables were set to missing for the cases where the person self-reported as non-Latino and other race only. The missing values were imputed through a series of hot-deck imputations where pools of donors were created by matching the structure of the household and non-missing values of race and ethnicity of the adult, child, or adolescent in the household within geographic areas (i.e., stratum, region, or urban/rural area). For cases where there was no pool of donors based on household structure, missing values were imputed using the value of SRW2, SRAA2, SRAI2, SRAS2, and SRPI2 from another member of the household. Next, the variable OMBSRREO was assigned for the records with SRH=2 and SRO=1 using the imputed of values SRW2, SRAA2, SRAI2, SRAS2, and SRPI2. Table 8-8 shows the counts and percentages of imputed OMBSRREO values by type of extended interview (adult, child, and adolescent).

Table 8-8. Number and percentage of completed interviews with missing OMB race and ethnicity by extended interview type

OMB Race-ethnicity (OMBSRREO)	Total		Extended interview type: Adult		Child		Adolescent	
	Imputed Count	%	Imputed Count	%	Imputed Count	%	Imputed Count	%
1. Latino	NA		NA		NA		NA	
2. Non-Latino White alone	85	0.13%	76	0.12%	7	0.01%	2	0.00%
3. Non-Latino African American alone	12	0.02%	11	0.02%	1	0.00%	0	0.00%
4. Non-Latino Asian alone	2	0.00%	2	0.00%	0	0.00%	0	0.00%
5. Non-Latino American Indian/ Alaska Native alone	5	0.01%	5	0.01%	0	0.00%	0	0.00%
6. Non-Latino Native Hawaiian and Other Pacific Islander alone	0	0.00%	0	0.00%	0	0.00%	0	0.00%
7. Non-Latino two or more races	10	0.02%	9	0.01%	1	0.00%	0	0.00%
Total	114	0.18%	103	0.16%	9	0.01%	2	0.00%
Completed interviews	64,599		51,048		9,913		3,638	

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

8.4.3 Self-Reported Asian Ethnic Group

The person weights were raked using a dimension defined for Asian groups (Dimension 7). Since there was only one weight for the combined landline and supplemental list samples in CHIS 2007, we added a variable (OMBSRASO) for a raking dimension that would improve the estimates of the largest Asian ethnic groups in California. The variable OMBSRASO identifies the OMB non-Latino Asian ethnic group and is defined in Table 8-9.

Table 8-9. OMB Non-Latino Asian ethnic groups (OMBSRASO)

OMBSRASO	Description
1	Non-Latino Chinese alone
2	Non-Latino Korean alone
3	Non-Latino Filipino alone
4	Non-Latino Vietnamese alone
5	Other

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

The process to derive the variable OMBSRASO used the temporary OMB race variable SRAS2 previously created for the imputation of OMBSRREO. For records where SRAS2=1 (self-reported as OMB Asian alone or combined with some other race), five flags indicating the Asian ethnic groups of the respondent were derived using the Asian ethnic group

questions in the extended interview (questions AA5E_1 to AA5E_18 for adults, TI7_1 to TI 7_18 for adolescents, and CH7_1 to CH7_18 for children). The name and description of the Asian ethnic group flags are shown in Table 8-10.

Table 8-10. OMB Asian group flags

Variable	Description
SRCH	Self-reported Chinese
SRPH	Self-reported Filipino
SRKR	Self-reported Korean
SRVT	Self-reported Vietnamese
SRASO	Self-reported Other Asian ethnic group

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

The missing values of the OMB Asian group variables (SRCH, SRPH, SRKR, SRVT, and SRASO) were imputed in the same way as the OMB race variables (or the temporary OMB race variables). A series of hot-deck imputations were run where pools of donors were created by matching the structure of the household and non-missing values of race, ethnicity, and Asian ethnic group of the adult, child, or adolescent in the household within geographic areas (i.e., stratum, region, or urban/rural area). For cases where there was no pool of donors based on household structure, race, ethnicity and Asian ethnic group, missing values were imputed using the values of SRCH, SRPH, SRKR, SRVT, and SRASO from another member of the household. The variable OMBSRASO was then created using the variables SRH, SRAA2, SRAI2, SRAS2, SRPI2, and the variables SRCH, SRPH, SRKR, SRVT, and SRASO after imputation. Table 8-11 shows the counts and percentages of imputed OMBSRASO values by type of extended interview (adult, child, and adolescent).

Table 8-11. Number and percentage of completed interviews with imputed OMB Asian ethnic group by extended interview type

			Extended interview type					
	Total		Adult		Child		Adolescent	
OMB Asian group (OMBSRASO)	Imputed count	%	Imputed count	%	Imputed count	%	Imputed count	%
1. Non-Latino Chinese	12	0.02	10	0.02	2	0.00	0	0.00
2. Non-Latino Korean	3	0.00	3	0.00	0	0.00	0	0.00
3. Non-Latino Filipino	4	0.01	3	0.00	0	0.00	1	0.00
4. Non-Latino Vietnamese	3	0.00	2	0.00	1	0.00	0	0.00
5. Other	1975	3.06	1262	1.95	479	0.74	234	0.36
Total	1997	3.09	1280	1.98	482	0.75	235	0.36
Completed interviews	64,599		51,048		9,913		3,638	

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

8.5 Self-Reported County and Self-Reported Stratum

In CHIS 2007, the geographic location variables such as self-reported county of residence, Los Angeles SPA, San Diego County Health Region, Census tract, and self-reported stratum were assigned after geocoding the geographic information collected during the interview (address of respondent, nearest street intersection, or self-reported county) or attached to the sample telephone number (the mailing address or ZIP Code covered by the telephone exchange). Table 8-12 shows the variables used in the geocoding process.

Table 8-12. Variables used in geocoding

Variable	Description	Source
AH42	County of residence (self report)	Adult questionnaire
AO1ADDR	Confirmed/corrected street address	Adult questionnaire
AO1CITY	Confirmed/corrected city	Adult questionnaire
AO1ZIP	Confirmed/corrected ZIP Code	Adult questionnaire
AM7	ZIP Code (self report)	Adult questionnaire
AO2ADDR	Street address (self report)	Adult questionnaire
AO2CITY	City (self report)	Adult questionnaire
AM8	Street name of residence (self report)	Adult questionnaire
AM9	Street name of nearest cross street (self report)	Adult questionnaire
M_ADDR	Street address (matched to phone number prior to interview)	Address mailing vendor
M_CITY	City (matched to phone number prior to interview)	Address mailing vendor
M_ZIP	ZIP Code (matched to phone number prior to interview)	Address mailing vendor
S_ZIP	ZIP Code (provided by sample vendor for every phone)	Sample vendor

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

The derived location variables SRSTRATA (self-reported stratum), SRCOUNTY (self-reported county), SR_LASPA (self-reported Los Angeles SPA), SR_HR (self-reported San Diego County Health Region) are household-level variables that were assigned to all adult, child and adolescent records within the same household before creating the raking dimensions. The variable SRSTRATA was used to create the cells for raking dimensions 1, 2, and 8 defined at the stratum or California region level while the variables SRCOUNTY (self-reported county), SR_LASPA (self-reported Los Angeles SPA), and SR_HR (self-reported San Diego County Health Region) were used to create the cells for raking dimension 4 defined for Los Angeles County and San Diego County.

Table 8-13 shows the distribution of adult respondents by self-reported stratum compared with the sampling stratum. Each stratum had migration in and migration out as a result of self-reports not matching the sampling stratum. This table shows that the net effect of cross-stratum migration is small, with the greatest differences for strata with the lowest geographic counts, as

indicated by the net agreement ratios (NAR) in the rightmost column of Table 8-13. The NAR is the number of respondents in the sampling stratum divided by the number of respondents in the self-reported stratum. A NAR value less than one indicates more in-migration than out-migration from the stratum, and a value greater than one the reverse. Most values are very close to one, indicating either very little migration or roughly equivalent rates of in- and out-migration.

Table 8-13. Distribution of self-reported strata and sampling strata

Stratum	Sampling stratum	Self-reported stratum	Net agreement ratio
Los Angeles	14,186	14,184	1.00
San Diego	6,234	6,237	1.00
Orange	3,701	3,637	1.02
Santa Clara	2,210	2,267	0.97
San Bernardino	2,265	2,282	0.99
Riverside	2,256	2,277	0.99
Alameda	2,042	1,939	1.05
Sacramento	1,828	1,821	1.00
Contra Costa	1,351	1,469	0.92
Fresno	1,035	1,045	0.99
San Francisco	1,090	1,072	1.02
Ventura	957	983	0.97
San Mateo	930	914	1.02
Kern	883	890	0.99
San Joaquin	781	777	1.01
Sonoma	722	733	0.98
Stanislaus	794	765	1.04
Santa Barbara	756	756	1.00
Solano	741	719	1.03
Tulare	793	794	1.00
Santa Cruz	743	724	1.03
Marin	694	697	1.00
San Luis Obispo	688	691	1.00
Placer	725	736	0.99
Merced	761	793	0.96
Butte	732	743	0.99
Shasta	719	750	0.96
Yolo	752	761	0.99
El Dorado	743	756	0.98
Imperial	802	795	1.01

Table 8-13. Distribution of self-reported strata and sampling strata (continued)

Stratum	Sampling stratum	Self-reported stratum	Net agreement ratio
Napa	699	721	0.97
Kings	780	781	1.00
Madera	764	755	1.01
Monterey	737	792	0.93
Humboldt	748	769	0.97
Nevada	701	687	1.02
Mendocino	744	715	1.04
Sutter	735	730	1.01
Yuba	770	726	1.06
Lake	693	677	1.02
San Benito	781	735	1.06
Colusa, Glenn, Tehama	604	577	1.05
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	569	569	1.00
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	554	552	1.00

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

9. VARIANCE ESTIMATION

This chapter describes the methods and results of computing sampling errors for CHIS 2007 data. The first section gives an overview of the reason for computing sampling errors and summarizes the precision of estimates for adults, children, and adolescents produced from three set of weights: (1) landline and list samples, (2) landline, list, and cell phone samples, and (3) area sample. The remainder of the chapter describes the methodology for producing estimates of sampling variability. Section 9.2 is a general review of the two main methods of computing sampling errors or variances of estimates from surveys with complex sample designs like CHIS 2007. Section 9.3 describes a replication method of variance estimation that can be used with the data. Section 9.4 shows how analysts can compute sampling errors for CHIS 2007 estimates using commercially available software.

9.1 Design Effects

To evaluate the precision of sample estimates derived from a survey, sampling errors are computed. Estimates of sampling errors can be used to make inferences about the size of the difference between two population parameters based on the values of corresponding sample estimates, their estimated precision, and the expected probability distribution of such a difference. Suppose an analyst wishes to compare the proportion of employed persons whose employer offers health care benefits in two counties in California. By taking the estimated sampling error of this difference into account, the analyst can make inferences about the size of the difference.

Inferences of this nature require an estimate of the precision or sampling error of the characteristic being investigated. There is a variety of ways of reporting the estimated precision of a survey estimate including:

- A standard error (the standard deviation of the estimate);
- A variance of an estimate (the standard error squared);
- A coefficient of variation (the ratio of the standard error to the estimate); or
- A confidence interval (the estimate plus or minus a multiple of the standard error).

Another way of describing the variability of an estimate from a survey is by using the "design effect." The concept of a design effect was introduced and popularized by Kish (1965) to account for the additional variability associated with complex sample designs involving stratification and clustering. The design effect is the ratio of the variance of the sample estimate for the survey (with its particular sample design and estimation method) to the variance of a simple random sample of the same sample size.

For a specific sample, the design effect, or *DEFF*, for an estimate from a survey can be estimated as

$$DEFF = \frac{\text{sampling variance of a complex sample}}{\text{sampling variance of a simple random sample}}.$$

At the analysis stage, the *DEFF* is useful because many procedures in statistical software assume the data are from a simple random sample when computing sampling errors of estimates. The *DEFF* can, in some circumstances, indicate the appropriateness of this assumption and can be used to adjust the sampling errors of the estimates to produce ones that are closer to the actual sampling errors (Skinner, Holt, and Smith, 1989).

Calculating the design effect for a proportion is straightforward because the variance of an estimated proportion in a simple random sample can be estimated easily. In this case, the estimated *DEFF* for a proportion is

$$DEFF_{PROP} = \frac{v(\hat{p})_{COMPLEX}}{v(\hat{p})_{SRS}},$$

where $\hat{p}$ is the estimated proportion, $v(\hat{p})_{SRS}$ is the variance estimate of the estimated proportion assuming a simple random sample, and $v(\hat{p})_{COMPLEX}$ is the variance of the estimated proportion accounting for the complex sample survey design.

In most surveys, design effects are larger than one. In CHIS 2007, design effects are greater than one mainly because the cases have different estimation weights (Kish, 1992). As will be seen shortly, design effects from the survey are considerably greater than one for some statewide estimates.

Design effects are of primary interest to data users. They reveal that the complex sample design and estimation procedures used resulted in estimates of variances that are greater

than what would be obtained from a simple random sample. A simple random sample design was not considered for CHIS 2007, because it would not have achieved the sample sizes for the specific domains of interest, in particular at the county/stratum level, for given resources. The design effects calculated from the CHIS 2007 data indicate that the sample design used in the survey needs to be taken into account when analyzing the data.

In CHIS 2007, as in most large-scale surveys, a large number of data items are collected. Each resulting variable has its own design effect. One way to summarize the design effects for the items is to compute *DEFF*s for a number of items and then average them. This average represents the design effects for similar items from the survey, as described in Wolter (1985).

The *DEFT* is the square root of the design effect, and it is similar to the *DEFF* but on the scale of the standard error of the estimate rather than the variance. Taking the square root of the *DEFF* has a smoothing affect on the variability.

The tables in the following sections show the *DEFF*s and *DEFT*s for selected items from either the adult, the child or the adolescent interviews. The *DEFT* is often considered a more convenient measure than the *DEFF*, because it can be used directly when computing confidence intervals for the estimates. See Verma, Scott, and O'Muircheartaigh (1980) for a discussion of the use of the *DEFT*. The main reason for presenting the *DEFT*s here is because it dampens some of the noise associated with the *DEFF*s. The maximum and minimum values of the *DEFF*s in the tables show that there is considerable variability in these quantities.

Before reviewing the tables in detail, it is important to discuss the most important factors that result in design effects larger than one. These factors are

- **Oversampling.** For the landline/list sample, the need for both county and state estimates required oversampling to produce stable estimates for these areas. This oversampling increased the design effect for statewide estimates. Another form of oversampling was the refusal subsampling, but this had a relatively minor effect on the design effects. The cell sample had minor disproportionate sampling and oversampling was not used in the area sample.

- **Within-Household Subsampling.** For all samples only one adult was selected in each household. In the landline/list sample one child and/or adolescent was sampled in each household. This subsampling contributed to the differential weights at the person level because persons in households with more persons

were subsampled at lower rates. In addition, young children (age 0 to 5 years) were sampled at twice the rate of older children (age 6 to 11 years)

- **Weighting Adjustments.** Differential weights were applied to reduce nonresponse bias and to make the estimates consistent with known population totals. The main reason for including these adjustments was to reduce biases in the estimates, but some of the adjustments may have increased the design effects for some estimates.

9.1.1 Design effect for the combined landline and list sample weights

Tables 9-1 to 9-3 present the *DEFF*s and *DEFT*s of the adult, child and adolescent interviews, respectively, for the landline/list sample. The first panel in the tables shows the average, median, minimum, and maximum *DEFF*s computed for a combination of categorical and continuous variables. The rightmost panel shows the average *DEFT* for the same items. The *DEFF*s and *DEFT*s were calculated using 15 items selected from the adult interview, 11 items from the child interview, and 17 items from the adolescent interview. All were calculated by stratum. The *DEFF*s and *DEFT*s for the adult, child, and adolescent are given separately for the categorical and the continuous variables in Appendix C.

Table 9-1 shows the average *DEFT*s for estimates of adult items are between 1.15 and 1.38 in most strata. This implies that for most strata the standard error of the estimates is about 15 to 38 percent greater than the expected standard error of a simple random sample. The average *DEFT* is very consistent by strata. The average *DEFT* for the state estimates is 1.37. This is larger than the county-level *DEFT*s as expected because counties were not sampled proportional to their population.

Table 9-1. Average *DEFF* and *DEFT* for estimates from the adult interview

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Total	1.93	1.82	3.28	0.78	1.37
Los Angeles	1.87	1.76	2.72	0.64	1.35
San Diego	1.68	1.61	2.32	0.67	1.28
Orange	1.93	1.83	3.99	0.64	1.37
Santa Clara	2.01	2.03	4.17	0.32	1.38
San Bernardino	1.49	1.42	2.74	0.43	1.20
Riverside	1.98	1.92	3.09	1.06	1.39
Alameda	1.62	1.64	2.61	0.72	1.26
Sacramento	1.44	1.44	1.86	0.55	1.19
Contra Costa	1.50	1.45	2.20	0.65	1.21
Fresno	1.60	1.57	2.87	0.68	1.25
San Francisco	1.60	1.73	2.26	0.31	1.25
Ventura	1.74	1.80	2.58	0.33	1.29
San Mateo	1.86	1.79	2.79	0.23	1.33
Kern	2.10	2.06	3.01	0.99	1.43
San Joaquin	1.70	1.62	3.12	0.62	1.28
Sonoma	1.62	1.56	2.95	0.15	1.24
Stanislaus	1.60	1.61	2.68	0.31	1.24
Santa Barbara	1.62	1.57	3.03	0.43	1.24
Solano	1.67	1.57	3.39	1.04	1.28
Tulare	1.40	1.26	2.13	0.70	1.17
Santa Cruz	1.48	1.43	2.13	0.66	1.20
Marin	1.57	1.57	2.46	0.65	1.23
San Luis Obispo	1.50	1.42	2.13	0.25	1.20
Placer	1.62	1.48	2.87	0.28	1.24
Merced	1.67	1.53	2.63	0.87	1.28
Butte	1.47	1.49	2.30	0.54	1.19
Shasta	1.64	1.56	3.17	0.64	1.26
Yolo	1.59	1.69	2.48	0.15	1.23
EL Dorado	1.41	1.38	2.07	0.25	1.17
Imperial	1.41	1.32	2.31	0.97	1.18
Napa	1.85	1.91	2.69	0.37	1.34
Kings	1.49	1.60	2.10	0.71	1.21
Madera	1.37	1.51	1.92	0.26	1.15
Monterey	1.68	1.72	2.25	0.89	1.29
Humboldt	1.89	1.43	6.05	0.81	1.32
Nevada	1.35	1.40	2.96	0.33	1.13
Mendocino	1.61	1.53	2.63	0.34	1.24
Sutter	1.60	1.48	2.30	0.48	1.25
Yuba	1.42	1.50	2.20	0.36	1.17
Lake	1.60	1.47	2.32	0.72	1.25
San Benito	1.54	1.65	2.81	0.58	1.22
Colusa, Glen, Tehama	1.59	1.60	2.38	0.39	1.24
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	1.60	1.58	2.35	0.64	1.25
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.34	1.34	1.86	0.43	1.15

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 9-2 shows the average *DEFT* for estimates from the child interview in each stratum for the landline/list weights. The average *DEFT* at the state level is 1.45. The average *DEFT*s for the counties vary between 1.10 and 1.40 for most strata; that is, the standard errors of these estimates are between 10 and 40 percent greater than expected from a simple random sample. The average *DEFT*s for estimates from the child interview are larger than those for the adult interview. This result is expected, because the subsampling at the person level for children is typically more variable than it is for adults (the number of children per household is more variable than the number of adults per household). Also, young children (age 0 to 5 years) were sampled at a rate twice that of older children (age 6 to 11 years) contributing further to the higher *DEFT*s. It is worth noting here that the *DEFT*s for the first four counties in Table 9-2 are larger than those of the counties in the rest of the table. This is a result of the differential sampling within these four counties between areas of high and low concentration of minorities. Of these four counties, San Diego County has the smallest (average *DEFT* = 1.37) because the large sample. When calculating the *DEFT*s, all samples were used, including the San Diego geographic oversample. This sample increased the number of children in San Diego, which increased the precision associated with child estimates.

Table 9-2. Average *DEFF* and *DEFT* for estimates from the child interview

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Total	2.13	2.02	2.99	1.13	1.45
Los Angeles	2.20	2.12	3.11	1.75	1.48
San Diego	1.63	1.54	2.36	1.15	1.27
Orange	1.71	1.47	3.14	1.06	1.29
Santa Clara	2.00	1.94	3.57	1.18	1.40
San Bernardino	1.60	1.64	2.26	0.37	1.25
Riverside	1.42	1.47	1.99	0.59	1.18
Alameda	1.40	1.41	1.99	0.74	1.18
Sacramento	1.31	1.32	1.80	0.99	1.14
Contra Costa	1.69	1.67	2.49	0.80	1.29
Fresno	1.53	1.40	2.59	1.04	1.22
San Francisco	1.44	1.42	1.87	1.05	1.19
Ventura	1.37	1.28	2.71	0.22	1.13
San Mateo	1.27	1.28	1.79	0.56	1.11
Kern	1.52	1.54	1.89	1.08	1.23
San Joaquin	1.79	1.48	4.08	0.64	1.30
Sonoma	1.28	1.15	2.07	1.04	1.12
Stanislaus	1.56	1.67	1.92	0.93	1.24
Santa Barbara	1.90	1.74	4.26	0.53	1.34
Solano	1.58	1.43	2.89	0.73	1.23
Tulare	1.36	1.37	1.74	1.07	1.16

Table 9-2. Average *DEFF* and *DEFT* for estimates from the child interview (continued)

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Santa Cruz	1.95	1.81	3.06	0.95	1.38
Marin	1.51	1.49	2.27	0.97	1.22
San Luis Obispo	1.55	1.12	2.85	0.44	1.20
Placer	1.53	1.52	2.03	1.13	1.23
Merced	1.70	1.86	2.10	0.24	1.27
Butte	1.23	1.23	1.53	0.87	1.11
Shasta	1.31	1.40	1.59	0.80	1.14
Yolo	1.52	1.49	2.27	0.98	1.22
EL Dorado	1.30	1.37	1.78	0.64	1.13
Imperial	1.51	1.59	1.84	0.77	1.22
Napa	1.47	1.18	2.20	0.89	1.20
Kings	1.45	1.33	2.00	1.05	1.20
Madera	1.67	1.71	2.16	0.96	1.29
Monterey	1.59	1.53	3.23	0.49	1.23
Humboldt	1.18	1.22	1.60	0.60	1.08
Nevada	1.16	1.15	1.77	0.46	1.06
Mendocino	1.30	1.31	1.91	0.54	1.12
Sutter	1.32	1.24	2.04	1.03	1.14
Yuba	1.89	1.74	3.22	0.82	1.36
Lake	1.56	1.53	1.86	1.30	1.25
San Benito	1.59	1.63	2.09	0.67	1.25
Colusa, Glen, Tehama	1.08	1.19	1.44	0.43	1.03
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	1.51	1.68	2.04	0.71	1.22
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.23	1.30	1.56	0.98	1.10

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

Table 9-3 shows that the average *DEFT* for items from the adolescent interviews are similar to those from the child interviews. Since the sampling for adolescents is similar to that for children, we expect a close correspondence between the two. The average *DEFT*s are slightly lower for adolescents than for children, primarily because there was no oversampling of adolescents by age and there are fewer adolescents than children per household, both of which reduce the variability in the weights. The average *DEFT* for the state estimates is 1.34. For most of the strata, the average *DEFT*s are between 1.05 and 1.32.

Table 9-3. Average *DEFF* and *DEFT* for estimates from the adolescent interview

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Total	1.80	1.84	2.38	1.48	1.34
Los Angeles	1.66	1.63	2.15	1.00	1.28
San Diego	1.23	1.23	1.65	0.78	1.10
Orange	1.50	1.41	2.21	0.83	1.22
Santa Clara	1.79	1.65	3.17	1.14	1.32
San Bernardino	1.54	1.52	2.55	0.83	1.23
Riverside	1.54	1.56	1.95	0.65	1.23
Alameda	1.63	1.38	3.09	0.77	1.25
Sacramento	1.53	1.50	2.18	1.03	1.23
Contra Costa	1.36	1.36	1.84	0.95	1.16
Fresno	1.23	1.24	1.86	0.82	1.10
San Francisco	2.02	2.09	4.29	0.33	1.37
Ventura	1.16	1.13	1.72	0.89	1.07
San Mateo	1.36	1.41	1.88	0.37	1.15
Kern	1.32	1.29	2.03	0.16	1.12
San Joaquin	1.51	1.61	1.99	0.83	1.22
Sonoma	1.11	1.10	1.40	0.70	1.05
Stanislaus	1.53	1.51	2.57	0.34	1.21
Santa Barbara	1.40	1.52	1.97	0.81	1.17
Solano	1.42	1.42	1.90	0.84	1.18
Tulare	1.07	1.19	1.81	0.41	1.02
Santa Cruz	1.69	1.68	2.91	0.45	1.26
Marin	1.34	1.32	1.80	0.95	1.15
San Luis Obispo	1.23	1.29	1.74	0.76	1.10
Placer	1.31	1.31	1.77	0.88	1.14
Merced	1.51	1.59	2.05	0.62	1.21
Butte	1.31	1.37	1.83	0.75	1.13
Shasta	1.43	1.40	2.66	0.72	1.18
Yolo	1.50	1.27	2.67	0.19	1.19
EL Dorado	1.32	1.23	2.05	1.00	1.14
Imperial	1.39	1.30	2.60	0.62	1.16
Napa	1.16	1.05	2.06	0.71	1.06
Kings	1.42	1.41	2.18	0.75	1.18
Madera	1.55	1.64	2.26	0.41	1.23
Monterey	1.31	1.32	1.79	0.71	1.13
Humboldt	1.34	1.36	1.95	0.74	1.15
Nevada	1.20	1.23	1.44	0.88	1.09
Mendocino	1.38	1.46	2.66	0.37	1.14
Sutter	1.33	1.41	1.77	0.83	1.15
Yuba	1.45	1.48	2.26	0.49	1.18
Lake	1.70	1.74	2.52	0.96	1.30

Table 9-3. Average *DEFF* and *DEFT* for estimates from the adolescent interview (continued)

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
San Benito	1.48	1.51	2.30	0.78	1.20
Colusa, Glen, Tehama	1.25	1.17	1.69	0.48	1.11
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	1.18	1.24	1.43	0.47	1.08
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.07	1.05	1.34	0.79	1.03

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

9.1.2 Design effects for the combined landline/ list/ cell phone sample weights

Similarly to Table 9-1, Table 9-4 presents the *DEFF*s and *DEFT*s for estimates for the same items from the adult interviews for the combined landline/list/cell phone sample weights. Tables for children and teens are not presented because only adults were interviewed in the cell phone samples.

Table 9-4 shows the average *DEFT*s for stratum-level estimates of adult items are between 1.20 and 2.96 for most strata. This implies that for most strata the standard error of the estimates is about 20 to 196 percent greater than the expected standard error of a simple random sample. The average *DEFT* for the state estimates is 2.30. The *DEFFs* and *DEFTs* of the combined landline, list, and cell phone sample weights are larger than those of the landline/list sample weights because the samples were not selected at the same rates.

Table 9-4. Average *DEFF* and *DEFT* for estimates from the adult interview for combined landline/list/cell phone sample weights

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Total	1.43	2.47	3.86	0.14	2.30
Los Angeles					
Antelope Valley	1.09	1.17	3.61	0.13	1.47
San Fernando	1.27	1.46	3.95	0.25	1.73
San Gabriel	1.47	2.15	4.79	0.71	2.33
Metro	1.16	1.24	3.23	0.50	1.48
West	1.23	1.57	3.93	0.46	1.66
South	1.98	3.62	18.44	0.29	4.67
East	1.67	3.47	5.09	0.25	3.00
South Bay	1.33	1.66	4.82	0.54	1.88
San Diego	1.67	2.58	4.74	1.52	2.85

Table 9-4. Average *DEFF* and *DEFT* for estimates from the adult interview for combined landline/list/cell phone sample weights (Continued)

Stratum	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Orange	1.53	2.22	3.71	0.75	2.40
Santa Clara	1.31	2.06	3.73	0.00	2.02
San Bernardino	1.23	1.78	2.99	0.03	1.71
Riverside	1.34	2.14	3.22	0.04	2.01
Alameda	1.32	2.02	3.75	0.00	1.99
Sacramento	1.13	1.42	2.63	0.05	1.42
Contra Costa	1.17	1.67	2.81	0.00	1.59
Fresno	1.24	1.87	3.67	0.03	1.74
San Francisco	1.22	1.91	3.35	0.00	1.76
Ventura	1.29	2.14	3.12	0.03	1.88
San Mateo	1.29	2.05	3.31	0.03	1.93
Kern	1.25	1.95	2.68	0.03	1.75
San Joaquin	1.22	1.77	2.94	0.04	1.65
Sonoma	1.27	1.72	2.98	0.17	1.76
Stanislaus	1.22	1.98	2.97	0.05	1.71
Santa Barbara	1.26	2.25	3.09	0.02	1.85
Solano	1.47	2.30	6.62	0.07	2.46
Tulare	1.13	1.64	2.46	0.01	1.44
Santa Cruz	1.23	1.83	3.14	0.02	1.71
Marin	1.46	2.53	4.64	0.68	2.25
San Luis Obispo	1.25	1.72	3.53	0.14	1.75
Placer	1.26	1.44	4.72	0.24	1.79
Merced	1.48	1.71	16.39	0.08	2.85
Butte	1.13	1.26	3.27	0.00	1.48
Shasta	1.17	1.87	3.00	0.00	1.62
Yolo	1.48	2.39	8.27	0.11	2.52
EL Dorado	1.26	1.64	2.83	0.27	1.64
Imperial	1.23	1.55	5.24	0.23	1.71
Napa	1.27	2.00	3.54	0.16	1.79
Kings	1.12	1.51	2.54	0.06	1.40
Madera	1.12	1.70	2.06	0.04	1.41
Monterey	1.22	1.59	4.33	0.09	1.67
Humboldt	1.29	1.45	5.37	0.00	2.13
Nevada	1.04	1.26	2.25	0.00	1.20
Mendocino	1.74	2.52	15.56	0.04	3.78
Sutter	1.19	1.59	3.39	0.00	1.69
Yuba	1.12	1.29	2.33	0.36	1.33
Lake	1.18	1.39	3.31	0.09	1.57
San Benito	1.58	2.25	8.01	0.15	2.96
Colusa, Glen, Tehama	1.21	1.66	2.28	0.41	1.52
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	1.08	1.22	2.86	0.00	1.33
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.00	1.22	1.93	0.00	1.13

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

9.1.3 Design effect for estimates from the area sample weights

Similarly to the tables in previous sections, Table 9-5 shows the *DEFF*s and *DEFT*s of estimates from adult interviews computed using the area sample weights. The average *DEFT*s for estimates of adult items are between 1.01 to 1.33 across SPAs. This implies that for SPAs the standard error of the estimates is about 1 to 33 percent greater than the expected standard error of a simple random sample. The average *DEFT* for Los Angeles County area estimates is 1.16, which is the lowest of all sets of weights. There was no geographic oversampling for the area sample, so the DEFTs for overall estimates would be expected to be within the range of those for the SPA-level estimates.

Table 9-5. Average *DEFF* and *DEFT* for estimates from the area sample adult interview

Special Planning Area	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
Total – Los Angeles County	1.40	1.35	0.40	2.30	1.16
Antelope Valley	2.04	1.84	0.26	7.01	1.33
San Fernando	1.33	1.06	0.23	3.69	1.10
San Gabriel	1.67	1.85	0.67	2.58	1.27
Metro	1.09	0.93	0.50	3.27	1.01
West	1.25	1.23	0.52	2.19	1.09
South	1.20	1.21	0.27	2.19	1.07
East	1.32	1.43	0.26	2.13	1.12
South Bay	1.77	1.56	0.52	4.85	1.28

Source: UCLA Center for Health Policy Research, 2007 California Health Interview Survey.

9.2 Methods for Variance Estimation

Variance estimation procedures have been developed to account for the complex sample design. Using these procedures, factors such stratification, multistage sampling, and the use of differential sampling rates to oversample a targeted subpopulation can be appropriately reflected in estimates of sampling error. The two main methods are replication and linearization or the Taylor series approximation. Wolter (1985) is a useful reference on the theory and applications of these methods. Shao (1996) is a more recent review paper that compares these methods. The rest of this section briefly reviews these methods.

The basic idea behind replication is to draw subsamples from the sample, compute the estimate from each of the subsamples, and estimate the variance of the original sample using the variability of the subsample estimates. Specifically, subsamples of the original "full" sample are selected to calculate subsample estimates of a parameter for which a "full-sample" estimate of

interest has been generated. The variability of these subsample estimates about the estimate for the full sample can then be assessed. The subsamples are called replicates, and the estimates from the subsamples are called replicate estimates. Rust and Rao (1996) discuss balanced repeated replication (BRR) and jackknife replication, two general approaches to forming subsamples. They show how the units included in the subsample can be defined using variance strata and units. They also describe how these methods can be implemented using replicate weights.

Replicate weights are created to produce the corresponding replicate estimate. Each replicate weight is computed using the same estimation steps as the full sample weight but using only the subsample of cases comprising each replicate. Once the replicate weights are developed, it is a straightforward matter to compute variance estimates for sample estimates of interest. The variance estimate takes the following form:

$$v\left(\hat{\theta}\right) = c \sum_{k=1}^{G} \left(\hat{\theta}_{(k)} - \hat{\theta}\right)^2 \qquad (1)$$

where

θ is an arbitrary parameter of interest;
$\hat{\theta}$ is the estimate of θ based on the full sample;
$\hat{\theta}_{(k)}$ is the k^{th} estimate of θ based on the observations included in the k^{th} replicate;
G is the total number of replicates formed;
c is a constant that depends on the replication method; and
$v\left(\hat{\theta}\right)$ is the estimated variance of θ.

The other widely used method for variance estimation for complex sample surveys is called linearization and is based on the Taylor series approximation. In this method, the Taylor series linearization of a statistic is formed and then substituted into the formula for calculating the variance of a linear estimate appropriate for the sample design. Linearization relies on the simplicity associated with estimating the variance for a linear statistic even with a complex sample design.

9.3 Design of Replicates

In CHIS 2007, a paired unit jackknife method (JK2[14]), a form of jackknife replication, was selected for computing variances. This section provides details on setting up the replication structure, including the definition of the variance strata and units.

Two major reasons for using replication to estimate variances for CHIS 2007 are operational convenience and the ability to reflect all components of the design and estimation in the estimates of variability. With respect to operational convenience, once replicate weights are constructed, it is very simple to compute estimates of sampling errors. No special care is needed for subgroups of interest, and no knowledge of the sample design is required. If an estimator is needed that was not previously considered, replication methods can be easily used to develop an appropriate estimate of variance. In such a case, variance estimates using a Taylor series approach would require additional work. The variance estimation stratum and unit must also be included in the file for the Taylor series method.

The second reason for using replication is probably more important. Both the nonresponse and raking types of adjustments made in developing the CHIS 2007 analysis weights affect the sampling errors of the estimates produced from the survey. Furthermore, the set of weights created in CHIS 2007 combined samples from different frames and were raked to the same control totals. The replicate weights prepared for CHIS reflect all such aspects of weighting and raking. Currently existing software for using the Taylor series method for variance estimation cannot reflect these weighting adjustments. In some Taylor series software poststratification can be taken into account, but only in specific situations.

In the JK2 replication method, adjacent pairs of sampled telephone numbers are treated as having been sampled from the same stratum. Each pair of sampled telephone numbers is treated as an implicit stratum, where each such stratum is defined by the sort order used in the sample selection of telephone numbers. In this method, the constant, c, in equation (1) equals 1. This approach has been used in previous cycles of CHIS and in other RDD studies such as the National Household Education Survey, an ongoing national RDD survey starting in 1991 and with a most recent cycle in 2007 (Hagedorn, et. al., 2008).

[14] This method is denoted as JK2 in the software program, WesVar, which was used to compute all the sampling errors in this report.

The first step in designing the replicate structure is to determine the number of variance estimation strata. In the JK2 method, the number of replicates is equal to the number of variance estimation strata. The choice of the number of variance estimation strata is based on the desire to obtain an adequate number of degrees of freedom to ensure stable estimates of variance while not having so many as to make the cost of computing variance estimates unnecessarily high. Generally, at least 30 degrees of freedom are needed to obtain relatively stable variance estimates. A number greater than 30 is often targeted because there are other factors that reduce the contribution of a replicate to the total number of degrees of freedom, especially for estimates of subgroups.

For CHIS 2007 and previous cycles of CHIS, we elected to create 80 variance estimation strata, even though many more could have been created. For the landline and cell phone samples, the 80 variance strata were formed as follows. First, the sampled telephone numbers were arranged in the same sort order used in sample selection. Next, adjacent sampled telephone numbers were paired to establish initial variance estimation strata (the first two sampled phone numbers were the first initial stratum, the third and fourth sampled telephone numbers were the second initial stratum, etc). Each telephone number in the pair was randomly assigned to be either the first or second variance unit within the variance stratum. Each pair was sequentially assigned to one of 80 final variance estimation strata (the first pair to variance estimation stratum 1, the second to stratum 2, ..., the 80th stratum pair to stratum 80, the 81st pair to stratum 1, etc.). As a result, each variance stratum had approximately the same number of telephone numbers. The same process was followed for each sampling stratum.

For the area sample, we used the same replication method and same number of replicates as for the combined samples to facilitate the computation of variance estimates. Addresses were sorted and paired when forming the variance strata.

Once the variance strata are created, the replicate weights can be created. The full replicate weights are constructed by first modifying the full sample base weights. The replicate base weight for replicate k for record i is

$$w_i^{(k)} = \begin{cases} 2w_i, & \text{if } i \text{ is in variance stratum } k \text{ and variance unit 1} \\ 0, & \text{if } i \text{ is in variance stratum } k \text{ and variance unit 2} \\ w_i, & \text{if } i \text{ is not in variance stratum } k \end{cases}$$

The same sequence of weighting adjustments used in the full sample weight is then applied to the replicate base weights to create the final replicate weights. Thus, all of the different components of the weighting process are fully reflected in the replicate weights, ranging from household adjustments (nonresponse, adjustment for household noncoverage, and adjustment to control totals) to person adjustments (nonresponse and raking).

9.4 Software for Computing Variances

In the past, most standard statistical software packages assumed a simple random sample when computing estimates of variance. As a result, estimates of variance from these packages had the potential to seriously understate the true variability of the survey estimates. However, in recent years, specialized commercial software has been developed to analyze data from complex surveys (Lepkowski and Bowles, 1996). In this section, we describe the elements needed to compute estimates for CHIS 2007 using some of these programs.

WesVar Version 4.3 (Westat, 2000) is a free software package developed and distributed by Westa**t.** WesVar uses replication methods to compute variance estimates. WesVar is an interactive program with a graphical interface that makes it simple to specify the estimates for sampling errors for estimates of interest. The data requests center on sessions called "workbooks." A workbook is a file linked to a specific WesVar data set. In a workbook, the user can request descriptive statistics, as well as analyze and create new statistics. Descriptive statistics of analysis variables are produced through "table requests." Regression requests support both linear and logistic regression models. Outputs include statistics of interest, such as the sum of weights, means, percentages, along with their corresponding standard errors, design effects, coefficients of variation, and confidence intervals.

To use WesVar with CHIS 2007 data, the only requirements are to identify the full and replicate weights that are on the data file and specify the replication method as JK2. This specification is made when a workbook is opened. All of the standard errors produced will properly account for the sample design and estimation methods because these features are accounted for in the replicate weights.

SUDAAN® (Research Triangle Institute, 2004) is a package developed by Research Triangle Institute to analyze data from complex sample surveys. SUDAAN is available as a standalone package or it can be called using SAS. SUDAAN and WesVar produce the same point

estimates. The difference between the two packages is in the method used to compute the variances. While WesVar uses replication exclusively, SUDAAN can use either a first-order Taylor series expansion approximation (linearization), or replication. When the Taylor series approximations are used, SUDAAN does not fully take into account complex weighting schemes such as nonresponse adjustments or raking, so the variance estimates will be different than estimates calculated using replication. On the other hand, if the user specifies replication as the variance estimation method, the estimates of variance computed in SUDAAN will take into account the sample design and weighting.

For descriptive statistics, SUDAAN offers two procedures: PROC CROSSTAB for categorical variables and PROC DESCRIPT for continuous variables. These procedures can be used to compute statistics of interest, such as sum of weights, means, and percentages along with their corresponding standard errors, design effects, and confidence intervals. Both procedures use the option DESIGN= to specify the type of survey design when calculating variance estimates. If no design type is specified using this option, then a standard "with replacement" design is assumed and linearization is used for variance estimation. Specifying JACKKNIFE assumes the use of replication. In this instance, the WEIGHT and NEST statements are also required. SUDAAN also contains procedures for computing other analytic statistics, such as those associated with linear and regression models. Consult the help manuals (available online) for more detail on the procedures and options available for SUDAAN.

Beginning in Version 9.1, SAS® has also included procedures to analyze survey data. Version 9.2 (SAS Institute, 2004. In Version 9.3, these procedures can use either the linearization or replication methods (include the REPWEIGHTS statement) to estimate the variance. The procedures in SAS for analyzing survey data are SURVEYMEANS, SURVEYREG, SURVEYFREQ and SURVEYLOGISTIC. The SURVEYMEANS procedure computes estimates of means, proportions, percentiles, and totals, Estimates of differences or other linear combinations are not available in SURVEYMEANS. The SURVEYFREQ procedure produces one-way tand cross tabulation tables for survey data. This procedure also computes estimates of odds ratios and relative risk estimates. The SURVEYREG procedure fits linear regression models while SURVEYLOGISTIC performs logistic regression for survey data and fit various links including the cumulative logit, generalized logit, probit, and complementary log-log functions.

Another software package that can be used to analyze survey data is STATA (version 10 is the latest version as of this writing) (STATA Corporation, 2008). STATA is a command driven, fully programmable statistical package used for managing, analyzing, and graphing data.

STATA was developed by StataCorp and is available for a variety of platforms, including DOS, Windows, Macintosh, and UNIX. STATA's statistical, graphical, and data management capabilities are fully expandable through programming.

STATA has a family of *svy*- commands to analyze data from sample surveys. The set of analytic methods in STATA is more exhaustive than any other package. The *svy* commands can be used to estimate a variety of quantities such as totals, proportions, means, linear combinations of means, and logistic regression parameters. Two-dimensional tables of totals and proportions, along with *DEFF*s for proportions can be produced using *svy* **tab.** The command *svy* **mean** can be used to produce the *DEFF*s for proportions by coding the analytical variable with values 0 and 1. To estimate totals using *svy* **total**, a variable with a value of 1 must be created for all records in the file. The *svy* command in the latest version of STATA can perform general linear modeling (**glm** command), nonlinear least squares estimation (**nl** command), and conditional logistic regression (**clogit** command) among others.

Like SUDAAN and SAS, STATA can use linearization (**linear** variance type option) or replication (**jack** variance type option) to estimate variances. Besides point estimates (proportions, means, ratios and totals) and their standard errors, STATA can compute confidence intervals, design effects, and misspecification effects. Design and misspecification effects are computed for means and proportions only.

When using linearization to estimate variances the software packages referred to above require auxiliary variables that provide information about the sample design. Two variables have been defined and included in the data files (TSVARSTR and TSVRUNIT). TSVARSTR is required for all analyses, but TSVRUNIT is required only when analyses are performed using the combined data. In other words, when separate analyses are done by adults, children or teens the variable TSVRUNIT is not required. The definitions of TSVARSTR and TSVRUNIT are

- **TSVARSTR** (Taylor's series variance stratum). The variable TSVARSTR indicates the variance strata to be used for software that computes estimates of variance using the Taylor series method. The variable TSVARSTR was created by sequentially numbering the sampling strata.

- **TSVRUNIT** (Taylor's series unit). The variable TSVRUNIT indicates the PSU. In this case the PSU is the sampled household. TSVRUNIT was created by sequentially numbering the PSU's within the sampling strata.

The same variables, TSVARSTR and TSVRUNIT, can be used for linearization variance estimation in SUDAAN, SAS, and STATA.

REFERENCES

Blumberg SJ, Luke JV.(2008). Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2008. National Center for Health Statistics. Available from: http://www.cdc.gov/nchs/nhis.htm. December 17, 2008.

Blumberg, S.J., Luke, J.V., Cynamon, M. L. (2006). Telephone coverage and health survey estimates: Evaluating the need for concern about wireless substitution. *American Journal of Public Health*, 96(5), 926-931.

Brackstone, G.J., and Rao, J.N.K. (1979). An investigation of raking ratio estimation. *Sankhya,* 41, 97-114.

Brick, J.M., Judkins, D., Montaquila, J., and Morganstein, D. (2002). Two-phase list-assisted RDD sampling. *Journal of Official Statistics*, 18(2), 203-215.

Brick, J.M., and Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5, 215-238.

Brick, J. M., Montaquila, J., Hagedorn, M. C., Roth, S. B., and Chapman, C. (2005). Implications for RDD design from an incentive experiment, *Journal of Official Statistics*, 21(4), 571-589.

Council of American Survey Research Organizations. (1982). *Special report: On the definition of response rates*. Port Jefferson, NY: CASRO.

Deville, J.C., and Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association,* 87, 376-382.

Elliott, M. R., Little, R.J.A., and Lewitzky, S. (2000). Subsampling callbacks to improve survey efficiency. *Journal of the American Statistical Association,* 95(451), 730-738.

Ford, B. M. (1983). *An overview of hot-deck procedures*: *Vol. 2. Incomplete data in sample surveys*. New York: Academic Press.

Ford, E.S. (1998). Characteristics of survey participants with and without a telephone: Findings from the third national health and nutrition examination survey. *Journal of Clinical Epidemiology*, 51, 55-60.

Groves, R.M. (1989). *Survey errors and survey costs*. New York: John Wiley and Sons.

Hagedorn, M., Roth, S.B., O'Donnell, K. Smith, S., and Mulligan, G. (2008). National Household Education Surveys Program of 2007: *Data File User's Manual, Volume I.* (NCES 2009-024). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hansen, M.H., and Hurwitz, W.N. (1946). The problem of nonresponse in sample surveys. *Journal of the American Statistical Association*, 41, 517–529.

Holt, D., and Smith, T.M.F. (1979). Post Stratification. *Journal of the Royal Statistical Society*. Series A (General), 142(1), 33-46.

Kalton, G., and Flores Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, 19(2), 81-97.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, 29, 119-127.

Kish, L. (1992). Weighting for unequal pi. *Journal of Official Statistic,* 8, 183-200.

Kish, L. (1965). *Survey sampling*. New York: John Wiley and Sons.

Lepkowski, J., and Bowles, J. (1996). Sampling error software for personal computers. *The Survey Statistician*, 35, 10-16.

Liu, B., Ferraro, D., Wilson, E., and Brick, J.M. (2004). Trimming extreme weights in household surveys. *Proceedings of the Survey Research Methods Section of the American Statistical Association* [CD-ROM]. Alexandria, VA: American Statistical Association.

Office of Management and Budget (1997). *Revisions to the standards for the classification of federal data on race and ethnicity*, 62(210).

Oh, H.L., and Scheuren, F.J. (1983). *Weighting adjustments for unit nonresponse*. Incomplete data in sample surveys: *Vol. 2.* W.G. Madow, I. Olkin, and D.B. Rubin (Eds.). Academic Press, 143-184.

Research Triangle Institute (2004). *SUDAAN user's manual, Release 9.0*. Research Triangle Park, NC: Research Triangle Institute.

Rizzo, Lou, Brick J.M., and Park, I. (2004). A minimally intrusive method for sampling persons in random-digit-dial surveys. *Public Opinion Quarterly,* 68, 267-274.

Rust, K.F., and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5, 282-310.

Sande, I. G. (1983). *Hot-deck imputation procedures, incomplete data in sample surveys*: *Vol. 3*. New York: Academic Press.

SAS Institute (2004). *SAS/STAT→ 9.1 User's Guide ,* Cary, NC: SAS Institute, Inc.

Shao, J. (1996). Resampling methods in sample surveys (with discussion). *Statistics*, 27, 203-254.

Skinner, C.J., Holt D., and Smith, T. M. F. (1989). *Analysis of complex surveys*. New York: John Wiley and Sons.

Stata Corporation (2005). *STATA survey data reference manual, Release 9*. College Station, TX: Stata Corporation.

State of California, Department of Finance (2004). *Population projections by race/ethnicity for California and it's counties, 2000-2050.* Sacramento, CA.

State of California, Department of Finance (2006a). *California county population estimates and components of change by year, July 1, 2000–2005*. Sacramento, CA.

State of California, Department of Finance (2006b). *Estimated race/ethnic population with age and sex detail*, 2000–2004. Sacramento, CA.

U.S. Census Bureau (2001). *2000 Summary File 1* (Prepared by the U.S. Census Bureau). California.

U.S. Census Bureau (2002). *Census 2000 Modified Race Data*. (Prepared by the U.S. Census Bureau).

U.S. Census Bureau (2003). *2002 American Community Survey PUMS file* (Prepared by the U.S. Census Bureau). California.

Verma, V., Scott, C., and O'Muircheartaigh, C. (1980). Sample designs and sampling errors for the world fertility survey. *Journal of the Royal Statistical Society*, A 143, 431-473.

Westat (2000). *WesVar ™ 4.0 user's guide.* Westat, Rockville, MD.

Wolter, K. (1985). *Introduction to variance estimation*. Springer-Verlag.

Appendix A

Table A-1. CHIS 2007 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean surname list and Korean and any other race but Vietnamese surname list)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Korean and any other race but Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
1.12	Los Angeles, San Fernando SPA – High Density	116,500	11,999	9.7108	893	85	5.1322	1659	0	0.0000
1.13	Los Angeles, San Gabriel SPA – High Density	364,800	11,099	32.8705	1709	155	8.0235	14039	13	32.7249
1.14	Los Angeles, Metro SPA – High Density	398,300	12,200	32.6984	3105	293	8.1070	4989	5	29.0058
1.17	Los Angeles, South SPA – High Density	113,200	8,998	12.6124	752	71	5.3333	1971	1	12.1667
1.18	Los Angeles, South Bay SPA – High Density	207,900	11,896	17.4559	722	64	6.8113	1727	2	18.3723
1.21	Los Angeles, Antelope Valley SPA – Low Density	211,100	6,600	32.0074	154	14	7.7000	325	0	0.0000
1.22	Los Angeles, San Fernando SPA – Low Density	1,654,700	31,395	52.7010	2,207	199	9.0451	4968	3	49.6800
1.23	Los Angeles, San Gabriel SPA – Low Density	937,300	24,992	37.5059	1343	117	8.8355	10128	6	37.5111
1.24	Los Angeles, Metro SPA – Low Density	651,700	23,597	27.6143	1261	123	7.2057	3562	0	0.0000
1.25	Los Angeles, West SPA – Low Density	1,048,400	17,999	58.2177	1012	90	9.5472	3100	0	0.0000
1.26	Los Angeles, South SPA –Low Density	683,200	13,100	52.1346	718	67	8.5476	1456	1	63.3043
1.27	Los Angeles, East SPA – Low Density	767,700	19,895	38.593	744	64	8.9639	2056	0	0.0000
1.28	Los Angeles, South Bay SPA – Low Density	1,066,400	28,193	37.8342	1464	136	8.6118	3486	3	35.2121
2.12	San Diego, North Central Health Region – High Density	201,300	4,940	40.8092	192	18	8.0000	904	0	0.0000
2.13	San Diego, Central Health Region – High Density	156,400	5,115	30.5951	81	10	6.7500	390	0	0.0000
2.21	San Diego, North Coastal Health Region – Low Density	400,400	13,098	30.5616	325	33	8.1250	809	0	0.0000
2.22	San Diego, North Central Health Region – Low Density	309,800	9,760	31.7429	293	26	10.1034	1042	1	30.6471
2.23	San Diego, Central Health Region – Low Density	336,600	14,085	23.9052	90	6	7.5000	368	0	0.0000
2.24	San Diego, South Health Region – Low Density	265,200	12,797	20.7408	128	10	5.8182	383	0	0.0000
2.25	San Diego, East Health Region – Low Density	385,900	11,900	32.4482	134	13	6.0909	296	0	0.0000
2.26	San Diego, North Inland Health Region – Low Density	360,400	11,500	31.3336	381	33	8.2826	1199	0	0.0000
3.1	Orange – High Density	956,200	32,491	29.419	3,431	315	8.1496	9020	4	27.4164
3.2	Orange – Low Density	1,681,300	26,400	63.6309	1,819	155	9.8324	5502	7	76.4167
4.1	Santa Clara - High	572,000	20,395	28.0848	783	63	8.7000	5343	4	27.2602
4.2	Santa Clara - Low	1,026,000	13,798	74.2944	1,881	160	9.6462	11244	10	73.4902
5	San Bernardino	1,267,300	23,798	53.2069	1,451	130	9.1835	4169	3	71.8793
6	Riverside	1,317,700	24,397	54.0098	1,152	111	8.5970	2801	5	48.2931
7	Alameda	1,525,600	28,200	54.0839	2,328	212	9.1654	14826	10	50.6007

Table A-1. CHIS 2007 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean surname list and Korean and any other race but Vietnamese surname list) (Continued)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Korean and any other race but Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
8	Sacramento	1,141,300	20,899	54.6097	1,037	100	8.3629	4855	2	55.8046
9	Contra Costa	931,000	16,197	57.4149	795	75	8.7363	3731	1	84.7955
10	Fresno	662,100	12,900	51.3033	562	57	8.6462	1726	2	47.9444
11	San Francisco	1,110,400	27,498	40.2989	1,815	177	8.7681	17021	9	41.6161
12	Ventura	618,000	11,099	55.6854	460	44	8.8462	1436	1	46.3226
13	San Mateo	816,500	15,300	53.3204	980	83	10.3158	5697	4	49.9737
14	Kern	529,400	9,299	56.9371	263	23	10.5200	569	0	0.0000
15	San Joaquin	409,200	8,398	48.7445	391	32	10.0256	1123	0	0.0000
16	Sonoma	450,300	7,299	61.6874	331	37	7.8810	663	0	0.0000
17	Stanislaus	318,900	6,800	46.9204	191	18	8.3043	441	0	0.0000
18	Santa Barbara	380,700	9,007	42.2882	234	19	8.6667	569	0	0.0000
19	Solano	300,100	7,799	38.4501	181	16	8.2273	655	1	50.3846
20	Tulare	282,800	8,799	32.127	83	6	11.8571	258	0	0.0000
21	Santa Cruz	271,000	8,000	33.8701	140	14	7.3684	408	0	0.0000
22	Marin	328,300	9,500	34.5522	223	16	10.6190	616	0	0.0000
23	San Luis Obispo	236,700	6,970	33.9384	128	7	12.8000	240	0	0.0000
24	Placer	299,400	7,148	41.8511	215	25	7.1667	557	0	0.0000
25	Merced	123,400	6,899	17.9041	90	7	5.6250	204	0	0.0000
26	Butte	160,700	5,397	29.7887	109	11	7.2667	264	0	0.0000
27	Shasta	139,800	5,300	26.3736	81	10	5.7857	123	0	0.0000
28	Yolo	140,700	6,399	21.9908	216	16	7.4483	976	1	19.1373
29	El Dorado	147,500	7,100	20.7745	103	11	5.7222	232	0	0.0000
30	Imperial	88,300	8,400	10.5191	47	1	5.2222	124	0	0.0000
31	Napa	119,200	8,699	13.701	66	5	6.0000	126	0	0.0000
32	Kings	70,700	6,598	10.7104	53	4	7.5714	79	0	0.0000
33	Madera	86,600	6,897	12.559	31	2	5.1667	83	0	0.0000
34	Monterey	350,800	11,199	31.3185	294	26	8.4000	620	0	0.0000
35	Humboldt	117,700	5,899	19.9354	71	6	8.8750	112	0	0.0000
36	Nevada	95,500	6,000	15.9277	75	9	4.1667	134	0	0.0000
37	Mendocino	89,600	6,500	13.7876	41	6	4.1000	96	0	0.0000

Table A-1. CHIS 2007 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean surname list and Korean and any other race but Vietnamese surname list) (Continued)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Korean and any other race but Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
38	Sutter	54,700	6,600	8.2783	41	2	8.2000	64	1	16.0000
39	Yuba	53,900	6,599	8.163	47	5	5.2222	88	0	0.0000
40	Lake	60,900	6,598	9.2276	55	6	4.5833	53	0	0.0000
41	San Benito	42,800	8,399	5.0959	15	0	0.0000	35	0	0.0000
42	Colusa, Glenn, Tehama	86,200	4,699	18.3486	34	2	6.8000	54	0	0.0000
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	180,900	6,400	28.2558	86	7	8.6000	149	0	0.0000
44	Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne	233,500	6,400	36.4861	112	6	12.4444	204	0	0.0000
	Total	30,512,800	804,531		40,249	3,664		156,147	100	

[1] Total number of possible phone numbers in eligible working 100 banks

[2] Realized number of sampled telephone numbers in strata.

Table A-2. CHIS 2007 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (Vietnamese, Vietnamese and any other race but Korean, Korean/Vietnamese surname list)

Sampling stratum	Description	Vietnamese surname list			Vietnamese and any other race but Korean surname list			Korean/Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
1.12	Los Angeles, San Fernando SPA – High Density	121	12	5.5000	72	2	9.0000	797	49	5.6929
1.13	Los Angeles, San Gabriel SPA – High Density	3906	498	6.2396	2129	23	24.7558	1788	128	9.8785
1.14	Los Angeles, Metro SPA – High Density	255	27	7.9688	183	4	20.3333	2128	162	8.8667
1.17	Los Angeles, South SPA – High Density	237	33	4.4717	138	2	10.6154	702	54	6.0517
1.18	Los Angeles, South Bay SPA – High Density	601	71	5.7789	164	1	20.5000	644	43	8.4737
1.21	Los Angeles, Antelope Valley SPA – Low Density	193	26	5.5143	39	0	0.0000	80	5	10.0000
1.22	Los Angeles, San Fernando SPA – Low Density	1814	227	7.0039	528	8	29.3333	1860	136	11.2727
1.23	Los Angeles, San Gabriel SPA – Low Density	3846	477	6.5969	1586	26	22.0278	1314	94	10.1077
1.24	Los Angeles, Metro SPA – Low Density	952	119	6.3046	465	6	23.2500	941	72	9.2255
1.25	Los Angeles, West SPA – Low Density	563	74	7.0375	359	3	44.8750	701	53	11.3065
1.26	Los Angeles, South SPA –Low Density	272	35	6.9744	72	1	24.0000	534	48	9.3684
1.27	Los Angeles, East SPA – Low Density	504	62	7.0000	199	3	24.8750	607	36	10.4655
1.28	Los Angeles, South Bay SPA – Low Density	1642	210	6.5680	503	11	16.2258	1007	79	9.1545
2.12	San Diego, North Central Health Region – High Density	1621	199	6.6434	263	2	37.5714	320	24	9.4118
2.13	San Diego, Central Health Region – High Density	1032	131	6.1796	149	2	16.5556	174	15	8.7000
2.21	San Diego, North Coastal Health Region – Low Density	471	53	6.6338	104	1	20.8000	182	15	10.1111
2.22	San Diego, North Central Health Region – Low Density	249	29	6.2250	115	4	12.7778	195	13	9.2857
2.23	San Diego, Central Health Region – Low Density	372	51	5.7231	77	1	19.2500	73	6	5.6154
2.24	San Diego, South Health Region – Low Density	147	16	6.3913	33	0	0.0000	106	6	6.6250
2.25	San Diego, East Health Region – Low Density	328	43	5.9636	58	1	29.0000	58	4	8.2857
2.26	San Diego, North Inland Health Region – Low Density	707	90	6.5463	189	1	31.5000	291	17	10.0345
3.1	Orange – High Density	14903	1846	6.4071	2482	49	18.8030	4255	304	9.3107
3.2	Orange – Low Density	4498	542	7.4470	947	16	27.8529	1660	121	11.0667
4.1	Santa Clara - High	7111	901	6.0519	1506	30	16.9213	1403	103	9.1104
4.2	Santa Clara - Low	6565	829	7.2382	2008	31	36.5091	2070	157	11.1892
5	San Bernardino	2029	255	7.0451	587	8	29.3500	1094	83	9.9455
6	Riverside	2131	263	7.2730	480	10	22.8571	805	68	9.2529
7	Alameda	5036	635	7.0730	2616	42	26.9691	2128	166	10.6935

Table A-2. CHIS 2007 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (Vietnamese, Vietnamese and any other race but Korean, Korean/Vietnamese surname list) (Continued)

Sampling stratum	Description	Vietnamese surname list			Vietnamese and any other race but Korean surname list			Korean/Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
8	Sacramento	3785	461	7.2649	1102	19	23.9565	864	62	10.9367
9	Contra Costa	1185	147	6.9298	572	7	38.1333	566	45	10.2909
10	Fresno	934	116	7.0226	200	4	40.0000	200	11	12.5000
11	San Francisco	2967	369	6.9000	3056	43	26.3448	1419	107	10.2086
12	Ventura	709	90	6.9510	173	3	34.6000	288	26	9.6000
13	San Mateo	685	85	7.2105	934	14	32.2069	560	43	11.4286
14	Kern	359	46	6.5273	56	1	28.0000	118	11	8.4286
15	San Joaquin	1145	137	6.8976	387	7	29.7692	212	12	12.4706
16	Sonoma	478	53	8.2414	106	1	17.6667	116	12	8.9231
17	Stanislaus	323	42	6.7292	59	1	14.7500	66	5	8.2500
18	Santa Barbara	289	35	6.5682	69	1	23.0000	106	9	9.6364
19	Solano	394	47	6.9123	110	2	55.0000	118	7	10.7273
20	Tulare	128	17	6.7368	21	0	0.0000	32	2	10.6667
21	Santa Cruz	171	24	5.8966	48	2	16.0000	53	6	8.8333
22	Marin	305	40	6.2245	73	0	0.0000	79	6	9.8750
23	San Luis Obispo	174	22	6.9600	36	1	36.0000	49	4	9.8000
24	Placer	324	42	6.7500	68	1	68.0000	102	10	10.2000
25	Merced	115	15	5.0000	18	1	18.0000	15	2	3.0000
26	Butte	192	29	5.3333	53	1	17.6667	23	1	7.6667
27	Shasta	71	10	6.4545	8	0	0.0000	8	0	0.0000
28	Yolo	346	44	6.1786	132	0	0.0000	136	11	10.4615
29	El Dorado	106	10	7.5714	25	0	0.0000	41	7	5.1250
30	Imperial	36	2	5.1429	8	1	8.0000	27	2	4.5000
31	Napa	66	7	6.6000	14	0	0.0000	20	3	5.0000
32	Kings	44	4	6.2857	6	0	0.0000	10	0	0.0000
33	Madera	50	7	4.5455	10	0	0.0000	7	0	0.0000
34	Monterey	269	39	5.7234	69	1	34.5000	215	21	7.6786
35	Humboldt	70	6	10.0000	11	1	11.0000	9	1	9.0000
36	Nevada	66	9	5.0769	11	0	0.0000	7	0	0.0000
37	Mendocino	64	6	5.8182	14	0	0.0000	9	0	0.0000

Table A-2. CHIS 2007 landline telephone sample frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (Vietnamese, Vietnamese and any other race but Korean, Korean/Vietnamese surname list) (Continued)

Sampling stratum	Description	Vietnamese surname list			Vietnamese and any other race but Korean surname list			Korean/Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
38	Sutter	49	2	7.0000	25	0	0.0000	8	0	0.0000
39	Yuba	47	8	3.9167	10	0	0.0000	12	1	6.0000
40	Lake	35	5	5.0000	4	0	0.0000	4	0	0.0000
41	San Benito	15	2	3.0000	4	0	0.0000	6	0	0.0000
42	Colusa, Glenn, Tehama	53	9	4.4167	9	0	0.0000	3	1	3.0000
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	118	17	6.2105	7	0	0.0000	7	0	0.0000
44	Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne	155	17	7.0455	15	0	0.0000	14	5	2.8000
	Total	78,428	9,775		25,574	400		33,446	2,494	

[1] Total number of possible phone numbers in eligible working 100 banks

[2] Realized number of sampled telephone numbers in strata.

Table A-3. CHIS 2007 cell-phone sample frame size, sample sizes, and base weights by sampling stratum

Sampling strata	Number of cell phone numbers	Number of sampled cell phone numbers	Weight
Total	37,006,000	42,490	871
2001	1,654,000	1,833	902
2002	2,729,000	2,345	1,164
2003	2,008,000	2,397	838
2005	1,657,000	1,594	1,040
2006	1,326,000	1,331	996
2007	1,653,000	1,713	965
2008	1,234,000	2,645	467
2009	1,392,000	1,518	917
2010	1,414,000	1,532	923
2011	1,812,000	1,662	1,090
2012	1,461,000	1,426	1,025
2013	894,000	1,007	888
2014	1,169,000	1,201	973
2015	1,342,000	1,871	717
2016	2,127,000	1,959	1,086
2017	1,772,000	2,361	751
2018	1,760,000	2,153	817
2019	1,877,000	2,161	869
2020	695,000	1,032	673
2021	628,000	804	781
2022	1,651,000	1,824	905
2023	1,565,000	2,552	613
2024	958,000	1,174	816
2025	915,000	1,035	884
2026	1,313,000	1,360	965

Table A-4. CHIS 2007 Los Angeles County area sample frame size, sample sizes, and base weights by sampling stratum

Sampling stratum (Los Angeles Service planning Area)	Frame		Sample		Addresses		Average Base weight
	Number of clusters	Number of occupied households	Number of clusters	Number of occupied households in cluster*	Number of addresses	Number of occupied households	
Total	31,883	3,133,774	212	3,133,774	33,446	4,289	751
Antelope Valley	1,181	95,493	6	95,493	1,003	122	776
San Fernando	6,410	679,886	46	679,886	9,290	933	751
San Gabriel	5,718	524,625	36	524,625	4,446	733	737
Metro	3,493	414,707	28	414,707	5,176	601	757
West	2,532	280,146	18	280,146	2,765	373	806
South	2,938	255,884	18	255,884	2,693	373	732
East	3,974	357,461	24	357,461	2,727	480	761
South Bay	5,637	525,572	36	525,572	5,346	674	729

*Based on Census 2000 data

Appendix B

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum

		All strata	Los Angeles	San Diego	Orange	Santa Clara
1.	Base weight					
1.1	Sample size	820,964	226,395	84,069	62,250	36,481
1.2	Sum of weight	30,512,800	8,221,200	2,416,000	2,637,500	1,598,000
1.3	Coefficient of variation (CV)	44.39	38.85	20.04	44.79	55.16
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	24,836,144	6,664,918	1,954,606	2,160,671	1,343,895
	b. Household refused – selected for ref. conversion	4,631,059	1,409,370	341,321	375,546	150,774
	c. Household refused – not selected for ref. conversion	1,045,596	148,817	119,952	102,160	103,439
2.2	Sum of weights after adjustment					
	a. Household never refused	24,836,144	6,664,918	1,954,606	2,160,671	1,343,895
	b. Household refused – selected for ref. conversion	5,676,656	1,558,187	461,273	477,706	254,214
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	30,512,800	8,223,105	2,415,879	2,638,377	1,598,108
2.4	Sample size	784,153	219,976	79,139	58,239	33,009
2.5	Coefficient of variation	44.73	37.59	21.35	42.46	55.45
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	26,319,200	7,073,820	2,074,445	2,275,878	1,406,585
	b. Household refused – selected for ref. conversion	2,804,038	642,051	203,949	243,318	171,118
	c. Household refused – not selected for ref. conversion	1,389,562	507,234	137,486	119,181	20,406

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		All strata	Los Angeles	San Diego	Orange	Santa Clara
3.2	Sum of weights after adjustment					
	a. Household never refused	26,319,200	7,073,820	2,074,445	2,275,878	1,406,585
	b. Household refused – selected for ref. conversion	4,193,600	1,149,285	341,434	362,499	191,524
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	30,512,800	8,223,105	2,415,879	2,638,377	1,598,108
3.4	Sample size	752,936	207,236	75,666	56,125	32,794
3.5	Coefficient of variation	50.03	45.09	34.61	48.15	57.51
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	3,401,552	795,984	274,860	255,745	158,586
	b. Residential – nonrespondent	4,755,431	1,359,263	387,378	416,709	221,755
	c. Unknown residential status (NA, NM)	3,621,278	982,571	328,188	311,257	194,063
	d. Nonresidential	18,734,538	5,085,287	1,425,453	1,654,665	1,023,704
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	3,401,552	795,984	274,860	255,745	158,586
	b. Residential – nonrespondent	4,755,431	1,359,263	387,378	416,709	221,755
	c. (NA, NM)	1,432,363	373,590	129,287	114,696	71,391
4.3	Sum of weights before adjustment					
	a. Residential – respondent	3,388,631	793,277	273,769	255,251	158,017
	b. Residential – nonrespondent	6,163,955	1,728,625	514,915	531,009	292,049
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	9,552,586	2,521,902	788,684	786,259	450,067
4.4	Sample size	165,899	44,075	15,579	11,531	6,043
4.5	Coefficient of Variation	67.18	61.28	55.43	64.61	70.40

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	All strata	Los Angeles	San Diego	Orange	Santa Clara
5. Supplemental list-sample eligibility adjustment					
5.1 Sum of weights before adjustment					
a. RDD – household	9,473,369	2,500,197	785,002	768,507	438,502
b. List eligible household	14,998	5,034	727	3,262	2,152
c. List ineligible household	4,718	1,123	211	415	509
d. Household with unknown list eligibility	59,501	15,549	2,744	14,074	8,903
5.2 Sum of weights before adjustment					
a. RDD household	9,473,369	2,500,197	785,002	768,507	438,502
b. List eligible household	58,590	17,959	3,035	13,324	9,045
c. List ineligible household	20,628	4,801	957	1,692	2,184
d. Household with unknown list eligibility	0	0	0	0	0
5.3 Sum of weights	9,552,586	2,522,956	788,995	783,523	449,732
5.4 Sample size	162,759	43,272	15,411	10,736	5,554
5.5 Coefficient of variation (CV)	65.72	59.78	54.50	58.47	62.81
6. Screener nonresponse adjustment					
6.1 Sum of weights before adjustment					
a. Respondents	3,496,647	822,044	281,955	271,172	169,490
b. Nonrespondents	6,055,940	1,700,912	507,039	512,351	280,242
6.2 Sum of weights after adjustment					
a. Respondents	9,552,586	2,522,956	788,995	783,523	449,732
b. Nonrespondents	0	0	0	0	0
6.3 Sum of weights	9,552,586	2,522,956	788,995	783,523	449,732
6.4 Sample size	86,121	21,091	8,603	5,313	2,923
6.5 Coefficient of variation	52.66	44.52	25.78	42.49	54.52
7. Multiple telephone adjustment					
7.1 Sum of weights before adjustment	9,515,232	2,512,892	786,298	780,398	446,535
7.2 Overall adjustment factor	0.97	0.97	0.97	0.97	0.97
7.3 Sum of weights	9,249,429	2,432,216	763,828	755,800	434,788

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	All strata	Los Angeles	San Diego	Orange	Santa Clara
7.4 Sample size	85,724	20,979	8,568	5,282	2,891
7.5 Coefficient of variation	54.16	46.26	28.65	44.17	55.74
8. Duplicate number adjustment					
8.1 Sum of weights before adjustment					
a. Not a duplicate number	9,249,171	2,432,152	763,785	755,800	434,788
b. Duplicate number	258	64	42	0	0
8.2 Sum of weights after adjustment					
a. Not a duplicate number	9,249,429	2,432,216	763,828	755,800	434,788
b. Duplicate number	0	0	0	0	0
8.3 Sum of weights	9,249,429	2,432,216	763,828	755,800	434,788
8.4 Sample size	85,715	20,977	8,567	5,282	2,891
8.5 Coefficient of variation	54.15	46.25	28.64	44.17	55.74
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment					
a. Household with child 1st procedure	734,543	180,386	62,807	63,599	40,252
b. Household w/o child 1st proc.-Sect. G comp.	5,090,277	1,268,649	417,562	395,798	246,405
c. Household w/o child 1st proc.-Sect. G not comp.	3,424,608	983,180	283,458	296,403	148,131
9.2 Sum of weights after adjustment					
a. Household with child 1st procedure	734,543	180,386	62,807	63,599	40,252
b. Household w/o child 1st proc.-Sect. G comp.	8,514,886	2,251,830	701,021	692,201	394,536
c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3 Sum of weights	9,249,429	2,432,216	763,828	755,800	434,788
9.4 Number HHs w/child 1st procedure					
9.5 Number HHs w/o child 1st proc.–Sect. G comp					
9.6 Coefficient of variation	57.61	50.46	32.01	46.63	55.58

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
1.	Base weight					
1.1	Sample size	24,277	24,854	29,265	21,543	16,472
1.2	Sum of weight	1,267,300	1,317,700	1,525,600	1,141,300	931,000
1.3	Coefficient of variation (CV)	13.09	12.48	17.46	16.11	12.32
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	980,764	1,022,504	1,288,677	934,547	767,936
	b. Household refused – selected for ref. conversion	271,483	278,066	150,835	149,350	155,989
	c. Household refused – not selected for ref. conversion	14,980	16,858	86,119	57,642	6,670
2.2	Sum of weights after adjustment					
	a. Household never refused	980,764	1,022,504	1,288,677	934,547	767,936
	b. Household refused – selected for ref. conversion	286,463	294,924	236,953	206,993	162,659
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	1,267,227	1,317,428	1,525,631	1,141,540	930,596
2.4	Sample size	23,733	24,285	27,093	20,165	16,189
2.5	Coefficient of variation	8.67	8.16	19.99	16.75	8.01
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	1,054,650	1,098,621	1,350,219	989,065	806,583
	b. Household refused – selected for ref. conversion	143,807	123,370	144,415	113,350	68,694
	c. Household refused – not selected for ref. conversion	68,770	95,436	30,996	39,125	55,319

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
3.2	Sum of weights after adjustment					
	a. Household never refused	1,054,650	1,098,621	1,350,219	989,065	806,583
	b. Household refused – selected for ref. conversion	212,577	218,807	175,411	152,475	124,013
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	1,267,227	1,317,428	1,525,631	1,141,540	930,596
3.4	Sample size	22,507	22,606	26,722	19,641	15,263
3.5	Coefficient of variation	18.78	25.17	24.95	24.76	23.41
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	169,763	178,431	155,357	141,770	102,224
	b. Residential – nonrespondent	223,242	239,772	203,468	169,126	136,298
	c. Unknown residential status (NA, NM)	137,413	154,904	183,861	157,835	106,362
	d. Nonresidential	736,808	744,321	982,944	672,809	585,711
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	169,763	178,431	155,357	141,770	102,224
	b. Residential – nonrespondent	223,242	239,772	203,468	169,126	136,298
	c. (NA, NM)	55,126	63,520	71,327	61,541	42,953
4.3	Sum of weights before adjustment					
	a. Residential – respondent	169,444	178,215	154,924	141,060	102,109
	b. Residential – nonrespondent	277,495	302,310	272,986	229,812	178,399
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	446,939	480,525	427,911	370,872	280,508
4.4	Sample size	6,036	5,897	5,002	4,341	3,174
4.5	Coefficient of Variation	33.53	43.31	47.43	46.88	46.04

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
5.	Supplemental list-sample eligibility adjustment					
5.1	Sum of weights before adjustment					
	a. RDD – household	444,757	478,479	422,432	367,687	279,372
	b. List eligible household	540	384	1,059	477	87
	c. List ineligible household	115	206	429	255	200
	d. Household with unknown list eligibility	1,527	1,456	3,990	2,453	849
5.2	Sum of weights before adjustment					
	a. RDD household	444,757	478,479	422,432	367,687	279,372
	b. List eligible household	1,932	1,459	4,346	2,076	312
	c. List ineligible household	576	890	1,964	1,057	873
	d. Household with unknown list eligibility	0	0	0	0	0
5.3	Sum of weights	447,265	480,828	428,742	370,819	280,558
5.4	Sample size	5,964	5,827	4,813	4,214	3,134
5.5	Coefficient of variation (CV)	31.86	41.95	43.19	43.56	44.71
6.	Screener nonresponse adjustment					
6.1	Sum of weights before adjustment					
	a. Respondents	172,854	182,535	161,698	145,146	104,327
	b. Nonrespondents	274,411	298,294	267,044	225,673	176,231
6.2	Sum of weights after adjustment					
	a. Respondents	447,265	480,828	428,742	370,819	280,558
	b. Nonrespondents	0	0	0	0	0
6.3	Sum of weights	447,265	480,828	428,742	370,819	280,558
6.4	Sample size	3,085	3,194	2,601	2,390	1,733
6.5	Coefficient of variation	15.33	17.77	26.27	23.16	16.74
7.	Multiple telephone adjustment					
7.1	Sum of weights before adjustment	446,269	479,370	425,983	369,135	279,374
7.2	Overall adjustment factor	0.98	0.98	0.97	0.98	0.97
7.3	Sum of weights	438,090	469,655	412,776	362,450	270,136
7.4	Sample size	3,075	3,178	2,578	2,373	1,722
7.5	Coefficient of variation	17.80	20.35	29.28	25.07	21.28

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
8. Duplicate number adjustment					
8.1 Sum of weights before adjustment					
a. Not a duplicate number	438,090	469,655	412,776	362,450	270,136
b. Duplicate number	0	0	0	0	0
8.2 Sum of weights after adjustment					
a. Not a duplicate number	438,090	469,655	412,776	362,450	270,136
b. Duplicate number	0	0	0	0	0
8.3 Sum of weights	438,090	469,655	412,776	362,450	270,136
8.4 Sample size	3,075	3,178	2,578	2,373	1,722
8.5 Coefficient of variation	17.80	20.35	29.28	25.07	21.28
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment					
a. Household with child 1st procedure	43,282	39,151	33,087	23,115	24,442
b. Household w/o child 1st proc.-Sect. G comp.	230,546	249,220	243,504	217,100	157,479
c. Household w/o child 1st proc.-Sect. G not comp.	164,262	181,284	136,185	122,235	88,216
9.2 Sum of weights after adjustment					
a. Household with child 1st procedure	43,282	39,151	33,087	23,115	24,442
b. Household w/o child 1st proc.-Sect. G comp.	394,808	430,504	379,689	339,335	245,694
c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3 Sum of weights	438,090	469,655	412,776	362,450	270,136
9.4 Number HHs w/child 1st procedure					
9.5 Number HHs w/o child 1st proc.–Sect. G comp					
9.6 Coefficient of variation	25.68	28.43	32.14	26.21	26.12

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Fresno	San Francisco	Ventura	San Mateo	Kern
1.	Base weight					
1.1	Sample size	13,090	28,203	11,263	15,529	9,380
1.2	Sum of weight	662,100	1,110,400	618,000	816,500	529,400
1.3	Coefficient of variation (CV)	10.89	13.61	11.10	10.46	8.70
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	532,926	967,683	498,913	691,852	417,166
	b. Household refused – selected for ref. conversion	110,177	117,132	66,923	114,735	73,790
	c. Household refused – not selected for ref. conversion	18,892	25,176	52,211	9,674	38,433
2.2	Sum of weights after adjustment					
	a. Household never refused	532,926	967,683	498,913	691,852	417,166
	b. Household refused – selected for ref. conversion	129,069	142,308	119,134	124,409	112,222
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	661,995	1,109,992	618,047	816,261	529,388
2.4	Sample size	12,612	27,196	10,238	15,219	8,659
2.5	Coefficient of variation	9.42	10.68	24.25	6.86	18.31
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	567,047	1,000,587	531,082	720,192	450,708
	b. Household refused – selected for ref. conversion	51,807	57,476	86,964	52,804	62,820
	c. Household refused – not selected for ref. conversion	43,141	51,929	0	43,265	15,859

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Fresno	San Francisco	Ventura	San Mateo	Kern
3.2	Sum of weights after adjustment					
	a. Household never refused	567,047	1,000,587	531,082	720,192	450,708
	b. Household refused – selected for ref. conversion	94,948	109,405	86,964	96,069	78,680
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	661,995	1,109,992	618,047	816,261	529,388
3.4	Sample size	11,894	26,125	10,238	14,465	8,475
3.5	Coefficient of variation	28.58	27.13	24.25	23.17	25.23
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	77,926	68,130	78,537	68,514	75,386
	b. Residential – nonrespondent	105,781	134,278	97,515	111,367	84,809
	c. Unknown residential status (NA, NM)	72,927	160,839	59,882	101,160	24,866
	d. Nonresidential	405,361	746,744	382,112	535,220	344,327
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	77,926	68,130	78,537	68,514	75,386
	b. Residential – nonrespondent	105,781	134,278	97,515	111,367	84,809
	c. (NA, NM)	31,309	57,528	24,459	40,151	11,369
4.3	Sum of weights before adjustment					
	a. Residential – respondent	77,823	67,888	78,203	68,141	75,159
	b. Residential – nonrespondent	136,822	190,324	121,444	151,162	95,699
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	214,645	258,212	199,647	219,303	170,858
4.4	Sample size	2,556	3,508	2,300	2,492	2,020
4.5	Coefficient of Variation	49.15	57.78	37.61	47.71	35.40

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Fresno	San Francisco	Ventura	San Mateo	Kern
5.	Supplemental list-sample eligibility adjustment					
5.1	Sum of weights before adjustment					
	a. RDD – household	213,862	254,428	199,001	218,116	170,515
	b. List eligible household	93	296	165	82	100
	c. List ineligible household	40	332	60	237	53
	d. Household with unknown list eligibility	650	3,156	422	868	190
5.2	Sum of weights before adjustment					
	a. RDD household	213,862	254,428	199,001	218,116	170,515
	b. List eligible household	344	1,198	718	267	345
	c. List ineligible household	206	1,558	234	1,037	258
	d. Household with unknown list eligibility	0	0	0	0	0
5.3	Sum of weights	214,411	257,185	199,953	219,420	171,119
5.4	Sample size	2,528	3,361	2,280	2,459	2,011
5.5	Coefficient of variation (CV)	48.29	54.47	36.24	46.54	34.52
6.	Screener nonresponse adjustment					
6.1	Sum of weights before adjustment					
	a. Respondents	78,907	71,252	80,156	69,925	77,138
	b. Nonrespondents	135,505	185,933	119,797	149,495	93,980
6.2	Sum of weights after adjustment					
	a. Respondents	214,411	257,185	199,953	219,420	171,119
	b. Nonrespondents	0	0	0	0	0
6.3	Sum of weights	214,411	257,185	199,953	219,420	171,119
6.4	Sample size	1,411	1,594	1,216	1,236	1,150
6.5	Coefficient of variation	20.87	22.09	30.10	17.76	25.19
7.	Multiple telephone adjustment					
7.1	Sum of weights before adjustment	213,919	255,278	199,040	218,048	170,721
7.2	Overall adjustment factor	0.99	0.96	0.97	0.96	0.98
7.3	Sum of weights	211,045	245,826	193,276	209,162	167,732
7.4	Sample size	1,408	1,575	1,209	1,225	1,145
7.5	Coefficient of variation	22.27	25.42	32.71	23.20	26.62

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	Fresno	San Francisco	Ventura	San Mateo	Kern
8. Duplicate number adjustment					
8.1 Sum of weights before adjustment					
a. Not a duplicate number	211,045	245,826	193,276	209,162	167,732
b. Duplicate number	0	0	0	0	0
8.2 Sum of weights after adjustment					
a. Not a duplicate number	211,045	245,826	193,276	209,162	167,732
b. Duplicate number	0	0	0	0	0
8.3 Sum of weights	211,045	245,826	193,276	209,162	167,732
8.4 Sample size	1,408	1,575	1,209	1,225	1,145
8.5 Coefficient of variation	22.27	25.42	32.71	23.20	26.62
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment					
a. Household with child 1st procedure	17,292	14,219	17,404	15,583	15,945
b. Household w/o child 1st proc.-Sect. G comp.	115,318	143,193	106,831	118,708	92,894
c. Household w/o child 1st proc.-Sect. G not comp.	78,436	88,414	69,041	74,871	58,894
9.2 Sum of weights after adjustment					
a. Household with child 1st procedure	17,292	14,219	17,404	15,583	15,945
b. Household w/o child 1st proc.-Sect. G comp.	193,754	231,607	175,872	193,579	151,788
c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3 Sum of weights	211,045	245,826	193,276	209,162	167,732
9.4 Number HHs w/child 1st procedure					
9.5 Number HHs w/o child 1st proc.–Sect. G comp					
9.6 Coefficient of variation	24.61	27.01	34.12	26.56	32.67

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
1.	Base weight					
1.1	Sample size	8,586	7,402	6,866	9,071	7,872
1.2	Sum of weight	409,200	450,300	318,900	380,700	300,100
1.3	Coefficient of variation (CV)	13.70	10.97	9.24	7.86	8.71
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	316,055	361,288	245,235	321,537	237,880
	b. Household refused – selected for ref. conversion	79,879	62,991	69,228	39,838	59,507
	c. Household refused – not selected for ref. conversion	13,002	25,741	4,411	19,291	2,538
2.2	Sum of weights after adjustment					
	a. Household never refused	316,055	361,288	245,235	321,537	237,880
	b. Household refused – selected for ref. conversion	92,881	88,732	73,639	59,128	62,045
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	408,936	450,020	318,874	380,665	299,924
2.4	Sample size	8,208	6,913	6,735	8,514	7,759
2.5	Coefficient of variation	10.86	15.11	6.62	15.34	5.69
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	341,591	383,728	266,603	340,421	252,133
	b. Household refused – selected for ref. conversion	39,839	45,484	28,575	40,244	47,792
	c. Household refused – not selected for ref. conversion	27,507	20,808	23,696	0	0

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
3.2	Sum of weights after adjustment					
	a. Household never refused	341,591	383,728	266,603	340,421	252,133
	b. Household refused – selected for ref. conversion	67,346	66,293	52,271	40,244	47,792
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	408,936	450,020	318,874	380,665	299,924
3.4	Sample size	7,719	6,672	6,258	8,514	7,759
3.5	Coefficient of variation	27.02	27.20	25.79	15.34	5.69
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	59,169	62,913	50,166	45,980	40,283
	b. Residential – nonrespondent	75,268	72,717	57,559	43,572	51,014
	c. Unknown residential status (NA, NM)	61,134	61,227	40,588	13,840	41,288
	d. Nonresidential	213,366	253,164	170,562	277,273	167,339
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	59,169	62,913	50,166	45,980	40,283
	b. Residential – nonrespondent	75,268	72,717	57,559	43,572	51,014
	c. (NA, NM)	26,548	27,582	18,111	6,070	18,097
4.3	Sum of weights before adjustment					
	a. Residential – respondent	59,169	62,234	50,166	45,600	40,207
	b. Residential – nonrespondent	101,812	98,338	75,661	49,222	69,036
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	160,981	160,572	125,826	94,822	109,242
4.4	Sample size	2,078	1,563	1,755	1,683	2,321
4.5	Coefficient of Variation	47.69	49.02	45.67	24.56	17.72

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
5.	Supplemental list-sample eligibility adjustment					
5.1	Sum of weights before adjustment					
	a. RDD – household	160,190	160,263	125,592	94,545	108,962
	b. List eligible household	97	0	0	35	93
	c. List ineligible household	58	54	21	36	14
	d. Household with unknown list eligibility	635	255	213	205	173
5.2	Sum of weights before adjustment					
	a. RDD household	160,190	160,263	125,592	94,545	108,962
	b. List eligible household	447	0	0	126	481
	c. List ineligible household	283	255	93	164	58
	d. Household with unknown list eligibility	0	0	0	0	0
5.3	Sum of weights	160,921	160,519	125,685	94,836	109,501
5.4	Sample size	2,045	1,550	1,743	1,673	2,313
5.5	Coefficient of variation (CV)	46.08	48.14	45.11	23.82	20.41
6.	Screener nonresponse adjustment					
6.1	Sum of weights before adjustment					
	a. Respondents	60,475	62,968	51,456	46,416	41,022
	b. Nonrespondents	100,445	97,551	74,229	48,419	68,480
6.2	Sum of weights after adjustment					
	a. Respondents	160,921	160,519	125,685	94,836	109,501
	b. Nonrespondents	0	0	0	0	0
6.3	Sum of weights	160,921	160,519	125,685	94,836	109,501
6.4	Sample size	1,146	914	1,019	957	1,035
6.5	Coefficient of variation	19.80	22.54	17.85	20.35	7.83
7.	Multiple telephone adjustment					
7.1	Sum of weights before adjustment	159,993	159,886	125,592	94,671	109,443
7.2	Overall adjustment factor	0.98	0.97	0.98	0.97	0.97
7.3	Sum of weights	157,262	154,386	123,058	92,257	106,681
7.4	Sample size	1,140	909	1,018	955	1,034
7.5	Coefficient of variation	21.70	26.23	20.49	22.94	13.66

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
8. Duplicate number adjustment					
8.1 Sum of weights before adjustment					
a. Not a duplicate number	157,262	154,386	123,058	92,257	106,681
b. Duplicate number	0	0	0	0	0
8.2 Sum of weights after adjustment					
a. Not a duplicate number	157,262	154,386	123,058	92,257	106,681
b. Duplicate number	0	0	0	0	0
8.3 Sum of weights	157,262	154,386	123,058	92,257	106,681
8.4 Sample size	1,140	909	1,018	955	1,034
8.5 Coefficient of variation	21.70	26.23	20.49	22.94	13.66
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment					
a. Household with child 1st procedure	12,728	11,509	12,975	7,532	8,174
b. Household w/o child 1st proc.-Sect. G comp.	79,853	92,409	65,803	54,833	56,491
c. Household w/o child 1st proc.-Sect. G not comp.	64,680	50,468	44,279	29,892	42,017
9.2 Sum of weights after adjustment					
a. Household with child 1st procedure	12,728	11,509	12,975	7,532	8,174
b. Household w/o child 1st proc.-Sect. G comp.	144,534	142,877	110,083	84,725	98,508
c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3 Sum of weights	157,262	154,386	123,058	92,257	106,681
9.4 Number HHs w/child 1st procedure					
9.5 Number HHs w/o child 1st proc.–Sect. G comp					
9.6 Coefficient of variation	28.59	28.77	28.31	29.30	21.38

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
1.	Base weight					
1.1	Sample size	8,824	8,046	9,562	7,004	7,226
1.2	Sum of weight	282,800	271,000	328,300	236,700	299,400
1.3	Coefficient of variation (CV)	4.26	6.66	7.03	5.64	9.19
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	229,001	226,469	275,294	198,436	236,679
	b. Household refused – selected for ref. conversion	50,407	41,171	33,613	23,074	41,202
	c. Household refused – not selected for ref. conversion	3,341	3,319	19,280	15,137	21,491
2.2	Sum of weights after adjustment					
	a. Household never refused	229,001	226,469	275,294	198,436	236,679
	b. Household refused – selected for ref. conversion	53,748	44,491	52,893	38,210	62,693
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	282,749	270,959	328,187	236,647	299,371
2.4	Sample size	8,704	7,921	8,966	6,538	6,666
2.5	Coefficient of variation	3.47	4.96	17.38	19.13	18.12
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	244,614	237,563	288,865	210,156	254,690
	b. Household refused – selected for ref. conversion	38,136	17,463	34,771	26,490	41,558
	c. Household refused – not selected for ref. conversion	0	15,933	4,551	0	3,123

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
3.2	Sum of weights after adjustment					
	a. Household never refused	244,614	237,563	288,865	210,156	254,690
	b. Household refused – selected for ref. conversion	38,136	33,396	39,322	26,490	44,681
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	282,749	270,959	328,187	236,647	299,371
3.4	Sample size	8,704	7,486	8,880	6,538	6,617
3.5	Coefficient of variation	3.47	25.09	20.88	19.13	20.26
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	35,720	33,337	34,821	35,483	45,077
	b. Residential – nonrespondent	41,533	36,202	41,898	28,586	45,571
	c. Unknown residential status (NA, NM)	21,575	34,899	33,780	13,250	38,294
	d. Nonresidential	183,922	166,521	217,688	159,328	170,430
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	35,720	33,337	34,821	35,483	45,077
	b. Residential – nonrespondent	41,533	36,202	41,898	28,586	45,571
	c. (NA, NM)	8,882	14,947	13,561	6,298	16,500
4.3	Sum of weights before adjustment					
	a. Residential – respondent	35,655	33,269	34,717	32,564	44,952
	b. Residential – nonrespondent	50,264	50,772	55,029	31,801	61,768
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	85,919	84,042	89,746	64,365	106,720
4.4	Sample size	2,305	1,534	1,621	1,339	1,650
4.5	Coefficient of Variation	12.09	52.43	38.90	32.26	36.86

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
5.	Supplemental list-sample eligibility adjustment					
5.1	Sum of weights before adjustment					
	a. RDD – household	85,875	83,884	89,561	64,193	106,503
	b. List eligible household	0	0	38	22	55
	c. List ineligible household	44	24	27	83	27
	d. Household with unknown list eligibility	0	134	121	67	134
5.2	Sum of weights before adjustment					
	a. RDD household	85,875	83,884	89,561	64,193	106,503
	b. List eligible household	0	0	170	128	195
	c. List ineligible household	153	111	75	401	140
	d. Household with unknown list eligibility	0	0	0	0	0
5.3	Sum of weights	86,029	83,995	89,805	64,722	106,839
5.4	Sample size	2,305	1,526	1,614	1,336	1,642
5.5	Coefficient of variation (CV)	12.40	52.00	38.37	33.52	36.14
6.	Screener nonresponse adjustment					
6.1	Sum of weights before adjustment					
	a. Respondents	36,303	33,843	35,292	33,362	45,570
	b. Nonrespondents	49,726	50,152	54,512	31,359	61,268
6.2	Sum of weights after adjustment					
	a. Respondents	86,029	83,995	89,805	64,722	106,839
	b. Nonrespondents	0	0	0	0	0
6.3	Sum of weights	86,029	83,995	89,805	64,722	106,839
6.4	Sample size	1,088	937	886	837	953
6.5	Coefficient of variation	6.50	19.22	23.96	26.68	22.81
7.	Multiple telephone adjustment					
7.1	Sum of weights before adjustment	85,801	83,884	89,425	64,109	106,420
7.2	Overall adjustment factor	0.98	0.95	0.96	0.97	0.98
7.3	Sum of weights	84,418	80,006	85,423	62,498	103,958
7.4	Sample size	1,085	935	883	831	949
7.5	Coefficient of variation	11.10	24.35	29.16	28.59	25.48

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
8. Duplicate number adjustment					
8.1 Sum of weights before adjustment					
a. Not a duplicate number	84,418	79,960	85,423	62,498	103,958
b. Duplicate number	0	46	0	0	0
8.2 Sum of weights after adjustment					
a. Not a duplicate number	84,418	80,006	85,444	62,498	103,958
b. Duplicate number	0	0	0	0	0
8.3 Sum of weights	84,418	80,006	85,444	62,498	103,958
8.4 Sample size	1,085	934	883	831	949
8.5 Coefficient of variation	11.10	24.23	29.14	28.59	25.48
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment					
a. Household with child 1st procedure	8,933	6,921	5,576	3,107	7,686
b. Household w/o child 1st proc.-Sect. G comp.	43,518	47,998	53,091	41,553	59,273
c. Household w/o child 1st proc.-Sect. G not comp.	31,967	25,087	26,777	17,837	36,999
9.2 Sum of weights after adjustment					
a. Household with child 1st procedure	8,933	6,921	5,576	3,107	7,686
b. Household w/o child 1st proc.-Sect. G comp.	75,484	73,085	79,867	59,390	96,272
c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3 Sum of weights	84,418	80,006	85,444	62,498	103,958
9.4 Number HHs w/child 1st procedure					
9.5 Number HHs w/o child 1st proc.–Sect. G comp					
9.6 Coefficient of variation	20.52	25.46	30.49	29.92	29.59

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Merced	Butte	Shasta	Yolo	El Dorado
1.	Base weight					
1.1	Sample size	6,924	5,439	5,320	6,471	7,128
1.2	Sum of weight	123,400	160,700	139,800	140,700	147,500
1.3	Coefficient of variation (CV)	5.76	8.05	5.28	8.46	5.26
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	96,326	124,863	109,868	114,488	115,227
	b. Household refused – selected for ref. conversion	24,353	25,128	16,700	15,365	18,833
	c. Household refused – not selected for ref. conversion	2,686	10,656	13,167	10,840	13,379
2.2	Sum of weights after adjustment					
	a. Household never refused	96,326	124,863	109,868	114,488	115,227
	b. Household refused – selected for ref. conversion	27,039	35,784	29,867	26,205	32,211
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	123,365	160,647	139,735	140,693	147,439
2.4	Sample size	6,758	5,051	4,804	5,937	6,464
2.5	Coefficient of variation	6.06	15.49	24.32	21.65	22.72
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	104,374	135,214	119,302	122,664	124,283
	b. Household refused – selected for ref. conversion	18,991	19,916	20,434	18,029	22,226
	c. Household refused – not selected for ref. conversion	0	5,517	0	0	930

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Merced	Butte	Shasta	Yolo	El Dorado
3.2	Sum of weights after adjustment					
	a. Household never refused	104,374	135,214	119,302	122,664	124,283
	b. Household refused – selected for ref. conversion	18,991	25,433	20,434	18,029	23,155
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	123,365	160,647	139,735	140,693	147,439
3.4	Sample size	6,758	4,921	4,804	5,937	6,438
3.5	Coefficient of variation	6.06	23.37	24.32	21.65	24.02
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	19,787	29,333	27,160	22,110	22,766
	b. Residential – nonrespondent	21,520	26,326	20,198	19,931	23,618
	c. Unknown residential status (NA, NM)	17,395	19,673	14,379	19,758	19,528
	d. Nonresidential	64,662	85,315	77,998	78,894	81,526
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	19,787	29,333	27,160	22,110	22,766
	b. Residential – nonrespondent	21,520	26,326	20,198	19,931	23,618
	c. (NA, NM)	8,182	9,634	6,859	8,185	9,126
4.3	Sum of weights before adjustment					
	a. Residential – respondent	19,769	29,303	27,133	22,066	22,704
	b. Residential – nonrespondent	29,711	35,925	27,058	28,053	32,664
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	49,480	65,228	54,192	50,119	55,368
4.4	Sample size	2,179	1,406	1,325	1,461	1,617
4.5	Coefficient of Variation	20.59	42.43	40.74	40.63	41.64

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Merced	Butte	Shasta	Yolo	El Dorado
5.	Supplemental list-sample eligibility adjustment					
5.1	Sum of weights before adjustment					
	a. RDD – household	49,413	65,154	54,176	49,940	55,340
	b. List eligible household	0	19	0	25	12
	c. List ineligible household	0	11	0	32	0
	d. Household with unknown list eligibility	67	44	16	123	16
5.2	Sum of weights before adjustment					
	a. RDD household	49,413	65,154	54,176	49,940	55,340
	b. List eligible household	0	51	0	104	38
	c. List ineligible household	0	55	0	147	0
	d. Household with unknown list eligibility	0	0	0	0	0
5.3	Sum of weights	49,413	65,261	54,176	50,191	55,377
5.4	Sample size	2,174	1,403	1,324	1,454	1,616
5.5	Coefficient of variation (CV)	20.48	42.17	40.70	40.48	41.57
6.	Screener nonresponse adjustment					
6.1	Sum of weights before adjustment					
	a. Respondents	20,098	29,532	27,189	22,525	22,935
	b. Nonrespondents	29,315	35,729	26,987	27,666	32,442
6.2	Sum of weights after adjustment					
	a. Respondents	49,413	65,261	54,176	50,191	55,377
	b. Nonrespondents	0	0	0	0	0
6.3	Sum of weights	49,413	65,261	54,176	50,191	55,377
6.4	Sample size	1,073	885	879	882	944
6.5	Coefficient of variation	7.75	20.26	28.18	26.15	26.95
7.	Multiple telephone adjustment					
7.1	Sum of weights before adjustment	49,413	65,141	54,176	50,044	55,377
7.2	Overall adjustment factor	0.98	0.98	0.99	0.98	0.97
7.3	Sum of weights	48,606	63,602	53,407	49,067	53,481
7.4	Sample size	1,073	883	879	880	944
7.5	Coefficient of variation	11.95	22.91	29.34	28.00	30.31

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	Merced	Butte	Shasta	Yolo	El Dorado
8. Duplicate number adjustment					
8.1 Sum of weights before adjustment					
a. Not a duplicate number	48,606	63,602	53,407	49,067	53,481
b. Duplicate number	0	0	0	0	0
8.2 Sum of weights after adjustment					
a. Not a duplicate number	48,606	63,602	53,407	49,067	53,481
b. Duplicate number	0	0	0	0	0
8.3 Sum of weights	48,606	63,602	53,407	49,067	53,481
8.4 Sample size	1,073	883	879	880	944
8.5 Coefficient of variation	11.95	22.91	29.34	28.00	30.31
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment					
a. Household with child 1st procedure	4,352	4,136	3,285	3,509	3,947
b. Household w/o child 1st proc.-Sect. G comp.	25,476	40,932	33,638	30,822	30,586
c. Household w/o child 1st proc.-Sect. G not comp.	18,778	18,534	16,484	14,737	18,947
9.2 Sum of weights after adjustment					
a. Household with child 1st procedure	4,352	4,136	3,285	3,509	3,947
b. Household w/o child 1st proc.-Sect. G comp.	44,254	59,466	50,122	45,558	49,533
c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3 Sum of weights	48,606	63,602	53,407	49,067	53,481
9.4 Number HHs w/child 1st procedure					
9.5 Number HHs w/o child 1st proc.–Sect. G comp					
9.6 Coefficient of variation	19.00	25.08	30.85	29.96	34.09

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Imperial	Napa	Kings	Madera	Monterey
1.	Base weight					
1.1	Sample size	8,406	8,714	6,606	6,906	11,286
1.2	Sum of weight	88,300	119,200	70,700	86,600	350,800
1.3	Coefficient of variation (CV)	2.70	2.99	1.65	3.08	7.67
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	68,373	97,158	53,917	66,240	292,637
	b. Household refused – selected for ref. conversion	18,240	20,140	15,826	19,364	48,180
	c. Household refused – not selected for ref. conversion	1,683	1,891	953	967	9,906
2.2	Sum of weights after adjustment					
	a. Household never refused	68,373	97,158	53,917	66,240	292,637
	b. Household refused – selected for ref. conversion	19,923	22,031	16,779	20,331	58,086
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	88,296	119,189	70,696	86,571	350,723
2.4	Sample size	8,242	8,567	6,512	6,821	10,915
2.5	Coefficient of variation	4.49	4.03	3.09	2.95	8.41
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	75,115	102,886	58,493	72,101	309,860
	b. Household refused – selected for ref. conversion	11,586	9,019	7,175	9,077	25,800
	c. Household refused – not selected for ref. conversion	1,595	7,283	5,028	5,393	15,063

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Imperial	Napa	Kings	Madera	Monterey
3.2	Sum of weights after adjustment					
	a. Household never refused	75,115	102,886	58,493	72,101	309,860
	b. Household refused – selected for ref. conversion	13,181	16,303	12,203	14,470	40,863
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	88,296	119,189	70,696	86,571	350,723
3.4	Sample size	8,103	8,081	6,069	6,411	10,516
3.5	Coefficient of variation	8.33	24.02	22.67	19.51	21.86
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	12,102	14,173	11,658	13,677	34,434
	b. Residential – nonrespondent	17,787	18,303	13,375	15,864	48,062
	c. Unknown residential status (NA, NM)	11,245	15,792	8,111	7,176	41,450
	d. Nonresidential	47,162	70,922	37,552	49,854	226,777
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	12,102	14,173	11,658	13,677	34,434
	b. Residential – nonrespondent	17,787	18,303	13,375	15,864	48,062
	c. (NA, NM)	4,817	6,452	4,005	3,279	15,335
4.3	Sum of weights before adjustment					
	a. Residential – respondent	12,091	13,995	11,637	13,664	34,403
	b. Residential – nonrespondent	22,604	24,470	17,370	19,023	63,224
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	34,695	38,465	29,006	32,687	97,627
4.4	Sample size	2,558	1,734	1,818	1,874	1,960
4.5	Coefficient of Variation	19.87	47.53	40.47	33.68	42.71

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Imperial	Napa	Kings	Madera	Monterey
5.	Supplemental list-sample eligibility adjustment					
5.1	Sum of weights before adjustment					
	a. RDD – household	34,695	38,380	28,970	32,673	97,380
	b. List eligible household	0	42	0	0	76
	c. List ineligible household	0	0	0	0	17
	d. Household with unknown list eligibility	0	43	37	15	152
5.2	Sum of weights before adjustment					
	a. RDD household	34,695	38,380	28,970	32,673	97,380
	b. List eligible household	0	154	0	0	226
	c. List ineligible household	0	0	0	0	56
	d. Household with unknown list eligibility	0	0	0	0	0
5.3	Sum of weights	34,695	38,534	28,970	32,673	97,663
5.4	Sample size	2,558	1,731	1,816	1,873	1,951
5.5	Coefficient of variation (CV)	19.87	47.68	40.49	33.69	42.25
6.	Screener nonresponse adjustment					
6.1	Sum of weights before adjustment					
	a. Respondents	12,232	14,222	11,762	13,839	35,365
	b. Nonrespondents	22,463	24,312	17,207	18,834	62,297
6.2	Sum of weights after adjustment					
	a. Respondents	34,695	38,534	28,970	32,673	97,663
	b. Nonrespondents	0	0	0	0	0
6.3	Sum of weights	34,695	38,534	28,970	32,673	97,663
6.4	Sample size	1,119	974	1,044	1,053	1,031
6.5	Coefficient of variation	7.41	15.24	14.39	12.55	17.03
7.	Multiple telephone adjustment					
7.1	Sum of weights before adjustment	34,665	38,453	28,944	32,673	97,422
7.2	Overall adjustment factor	0.98	0.97	0.99	0.98	0.97
7.3	Sum of weights	34,097	37,154	28,549	32,167	94,305
7.4	Sample size	1,118	972	1,043	1,053	1,028
7.5	Coefficient of variation	11.87	19.94	16.88	15.51	20.79

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	Imperial	Napa	Kings	Madera	Monterey
8. Duplicate number adjustment					
8.1 Sum of weights before adjustment					
a. Not a duplicate number	34,097	37,154	28,549	32,167	94,305
b. Duplicate number	0	0	0	0	0
8.2 Sum of weights after adjustment					
a. Not a duplicate number	34,097	37,154	28,549	32,167	94,305
b. Duplicate number	0	0	0	0	0
8.3 Sum of weights	34,097	37,154	28,549	32,167	94,305
8.4 Sample size	1,118	972	1,043	1,053	1,028
8.5 Coefficient of variation	11.87	19.94	16.88	15.51	20.79
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment					
a. Household with child 1st procedure	3,501	2,214	2,714	3,225	8,478
b. Household w/o child 1st proc.-Sect. G comp.	17,253	21,292	15,188	16,608	50,778
c. Household w/o child 1st proc.-Sect. G not comp.	13,344	13,648	10,646	12,334	35,049
9.2 Sum of weights after adjustment					
a. Household with child 1st procedure	3,501	2,214	2,714	3,225	8,478
b. Household w/o child 1st proc.-Sect. G comp.	30,597	34,940	25,835	28,942	85,827
c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3 Sum of weights	34,097	37,154	28,549	32,167	94,305
9.4 Number HHs w/child 1st procedure					
9.5 Number HHs w/o child 1st proc.–Sect. G comp					
9.6 Coefficient of variation	24.06	23.06	24.00	22.97	26.32

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Humboldt	Nevada	Mendocino	Sutter	Yuba
1.	Base weight					
1.1	Sample size	5,913	6,018	6,512	6,605	6,613
1.2	Sum of weight	117,700	95,500	89,600	54,700	53,900
1.3	Coefficient of variation (CV)	2.82	5.14	3.64	2.35	2.62
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	96,350	74,327	72,447	41,316	42,427
	b. Household refused – selected for ref. conversion	18,520	20,388	13,122	9,897	11,294
	c. Household refused – not selected for ref. conversion	2,811	733	4,018	3,485	180
2.2	Sum of weights after adjustment					
	a. Household never refused	96,350	74,327	72,447	41,316	42,427
	b. Household refused – selected for ref. conversion	21,331	21,121	17,140	13,382	11,474
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	117,681	95,448	89,587	54,698	53,901
2.4	Sample size	5,764	5,957	6,212	6,182	6,583
2.5	Coefficient of variation	5.63	3.77	10.82	13.35	2.22
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	102,046	79,015	76,393	44,705	45,464
	b. Household refused – selected for ref. conversion	9,390	8,046	8,222	7,196	8,436
	c. Household refused – not selected for ref. conversion	6,245	8,386	4,972	2,797	0

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Humboldt	Nevada	Mendocino	Sutter	Yuba
3.2	Sum of weights after adjustment					
	a. Household never refused	102,046	79,015	76,393	44,705	45,464
	b. Household refused – selected for ref. conversion	15,636	16,433	13,194	9,992	8,436
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	117,681	95,448	89,587	54,698	53,901
3.4	Sample size	5,492	5,448	5,935	5,932	6,583
3.5	Coefficient of variation	22.48	29.12	27.04	25.15	2.22
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	18,840	14,822	13,689	8,808	8,518
	b. Residential – nonrespondent	15,722	16,747	13,800	10,940	8,372
	c. Unknown residential status (NA, NM)	10,387	15,086	9,691	5,025	6,309
	d. Nonresidential	72,733	48,793	52,408	29,925	30,701
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	18,840	14,822	13,689	8,808	8,518
	b. Residential – nonrespondent	15,722	16,747	13,800	10,940	8,372
	c. (NA, NM)	5,033	7,291	4,214	2,266	2,771
4.3	Sum of weights before adjustment					
	a. Residential – respondent	18,840	14,822	13,689	8,800	8,510
	b. Residential – nonrespondent	20,749	23,962	17,988	13,154	11,130
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	39,589	38,784	31,677	21,954	19,640
4.4	Sample size	1,335	1,440	1,448	1,729	2,047
4.5	Coefficient of Variation	46.16	56.04	48.28	37.18	15.67

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Humboldt	Nevada	Mendocino	Sutter	Yuba
5.	Supplemental list-sample eligibility adjustment					
5.1	Sum of weights before adjustment					
	a. RDD – household	39,511	38,763	31,651	21,905	19,608
	b. List eligible household	20	0	0	0	0
	c. List ineligible household	34	21	0	0	0
	d. Household with unknown list eligibility	24	0	26	49	32
5.2	Sum of weights before adjustment					
	a. RDD household	39,511	38,763	31,651	21,905	19,608
	b. List eligible household	84	0	0	0	0
	c. List ineligible household	119	82	0	0	0
	d. Household with unknown list eligibility	0	0	0	0	0
5.3	Sum of weights	39,714	38,845	31,651	21,905	19,608
5.4	Sample size	1,334	1,440	1,446	1,726	2,044
5.5	Coefficient of variation (CV)	46.39	55.97	48.26	37.21	15.65
6.	Screener nonresponse adjustment					
6.1	Sum of weights before adjustment					
	a. Respondents	19,143	15,098	13,879	8,894	8,565
	b. Nonrespondents	20,571	23,747	17,772	13,011	11,043
6.2	Sum of weights after adjustment					
	a. Respondents	39,714	38,845	31,651	21,905	19,608
	b. Nonrespondents	0	0	0	0	0
6.3	Sum of weights	39,714	38,845	31,651	21,905	19,608
6.4	Sample size	900	897	919	957	1,038
6.5	Coefficient of variation	14.11	20.01	21.12	19.84	5.26
7.	Multiple telephone adjustment					
7.1	Sum of weights before adjustment	39,554	38,683	31,651	21,833	19,571
7.2	Overall adjustment factor	0.98	0.96	0.97	0.98	0.98
7.3	Sum of weights	38,607	37,329	30,749	21,470	19,202
7.4	Sample size	897	893	919	954	1,036
7.5	Coefficient of variation	17.89	24.12	24.13	21.74	11.05

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

	Humboldt	Nevada	Mendocino	Sutter	Yuba
8. Duplicate number adjustment					
8.1 Sum of weights before adjustment					
a. Not a duplicate number	38,607	37,264	30,749	21,470	19,202
b. Duplicate number	0	65	0	0	0
8.2 Sum of weights after adjustment					
a. Not a duplicate number	38,607	37,308	30,749	21,470	19,202
b. Duplicate number	0	0	0	0	0
8.3 Sum of weights	38,607	37,308	30,749	21,470	19,202
8.4 Sample size	897	890	919	954	1,036
8.5 Coefficient of variation	17.89	23.81	24.13	21.74	11.05
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment					
a. Household with child 1st procedure	2,689	2,522	1,533	1,713	1,455
b. Household w/o child 1st proc.-Sect. G comp.	25,098	23,284	19,974	12,524	10,445
c. Household w/o child 1st proc.-Sect. G not comp.	10,820	11,502	9,243	7,233	7,302
9.2 Sum of weights after adjustment					
a. Household with child 1st procedure	2,689	2,522	1,533	1,713	1,455
b. Household w/o child 1st proc.-Sect. G comp.	35,918	34,786	29,216	19,757	17,747
c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3 Sum of weights	38,607	37,308	30,749	21,470	19,202
9.4 Number HHs w/child 1st procedure					
9.5 Number HHs w/o child 1st proc.–Sect. G comp					
9.6 Coefficient of variation	20.44	24.95	25.35	23.67	18.85

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Lake	San Benito	Colusa, Glenn, Tehama	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1.	Base weight					
1.1	Sample size	6,609	8,401	4,711	6,424	6,428
1.2	Sum of weight	60,900	42,800	86,200	180,900	233,500
1.3	Coefficient of variation (CV)	2.61	1.00	4.52	5.06	5.93
2.	First refusal conversion subsampling adjustment					
2.1	Sum of weights before adjustment					
	a. Household never refused	47,815	34,252	67,064	154,265	191,850
	b. Household refused – selected for ref. conversion	6,824	8,554	14,007	15,823	24,100
	c. Household refused – not selected for ref. conversion	6,229	0	5,126	10,765	17,550
2.2	Sum of weights after adjustment					
	a. Household never refused	47,815	34,252	67,064	154,265	191,850
	b. Household refused – selected for ref. conversion	13,053	8,554	19,133	26,589	41,650
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
2.3	Sum of weights	60,868	42,806	86,197	180,854	233,500
2.4	Sample size	5,925	8,401	4,424	6,027	5,931
2.5	Coefficient of variation	27.21	0.83	13.41	18.91	21.57
3.	Second refusal conversion subsampling adjustment					
3.1	Sum of weights before adjustment					
	a. Household never refused	51,369	36,540	72,565	162,435	204,529
	b. Household refused – selected for ref. conversion	9,498	6,266	11,046	18,418	28,971
	c. Household refused – not selected for ref. conversion	0	0	2,585	0	0

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Lake	San Benito	Colusa, Glenn, Tehama	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
3.2	Sum of weights after adjustment					
	a. Household never refused	51,369	36,540	72,565	162,435	204,529
	b. Household refused – selected for ref. conversion	9,498	6,266	13,631	18,418	28,971
	c. Household refused – not selected for ref. conversion	0	0	0	0	0
3.3	Sum of weights	60,868	42,806	86,197	180,854	233,500
3.4	Sample size	5,925	8,401	4,320	6,027	5,931
3.5	Coefficient of variation	27.21	0.83	20.14	18.91	21.57
4.	Adjusting for unknown residential					
4.1	Sum of weights before adjustment					
	a. Residential – respondent	9,814	5,489	15,962	22,367	31,881
	b. Residential – nonrespondent	9,662	6,797	14,718	18,903	30,104
	c. Unknown residential status (NA, NM)	6,806	5,305	9,248	14,855	28,056
	d. Nonresidential	34,586	25,215	46,269	124,727	143,459
4.2	Sum of weights – allocating unknown residential					
	a. Residential – respondent	9,814	5,489	15,962	22,367	31,881
	b. Residential – nonrespondent	9,662	6,797	14,718	18,903	30,104
	c. (NA, NM)	3,614	2,094	4,487	6,309	13,568
4.3	Sum of weights before adjustment					
	a. Residential – respondent	9,814	5,489	15,962	22,283	31,844
	b. Residential – nonrespondent	13,250	8,883	19,190	25,191	43,604
	c. Estimated residential among unknown	0	0	0	0	0
4.3	Sum of weights	23,063	14,373	35,151	47,474	75,448
4.4	Sample size	1,479	2,410	1,313	1,112	1,251
4.5	Coefficient of Variation	44.00	13.61	36.67	37.27	42.33

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Lake	San Benito	Colusa, Glenn, Tehama	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
5.	Supplemental list-sample eligibility adjustment					
5.1	Sum of weights before adjustment					
	a. RDD – household	23,051	14,360	35,104	47,459	75,410
	b. List eligible household	0	6	0	0	0
	c. List ineligible household	0	0	18	0	14
	d. Household with unknown list eligibility	12	7	29	15	24
5.2	Sum of weights before adjustment					
	a. RDD household	23,051	14,360	35,104	47,459	75,410
	b. List eligible household	0	25	0	0	0
	c. List ineligible household	0	0	74	0	73
	d. Household with unknown list eligibility	0	0	0	0	0
5.3	Sum of weights	23,051	14,385	35,179	47,459	75,483
5.4	Sample size	1,478	2,409	1,311	1,111	1,249
5.5	Coefficient of variation (CV)	44.00	15.11	36.51	37.21	42.10
6.	Screener nonresponse adjustment					
6.1	Sum of weights before adjustment					
	a. Respondents	9,851	5,542	16,181	22,339	32,150
	b. Nonrespondents	13,200	8,843	18,998	25,120	43,333
6.2	Sum of weights after adjustment					
	a. Respondents	23,051	14,385	35,179	47,459	75,483
	b. Nonrespondents	0	0	0	0	0
6.3	Sum of weights	23,051	14,385	35,179	47,459	75,483
6.4	Sample size	899	1,077	805	689	744
6.5	Coefficient of variation	31.32	6.68	16.03	25.61	26.72
7.	Multiple telephone adjustment					
7.1	Sum of weights before adjustment	23,030	14,372	35,024	47,399	75,323
7.2	Overall adjustment factor	0.98	0.98	0.97	0.99	0.97
7.3	Sum of weights	22,583	14,081	34,062	46,754	73,431

Table B-1. Household weighting for the combined RDD, Korean and Vietnamese surname list samples by stratum (continued)

		Lake	San Benito	Colusa, Glenn, Tehama	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
7.4	Sample size	898	1,076	801	688	742
7.5	Coefficient of variation	33.22	12.23	19.69	27.44	29.28
8.	Duplicate number adjustment					
8.1	Sum of weights before adjustment					
	a. Not a duplicate number	22,562	14,081	34,042	46,754	73,431
	b. Duplicate number	20	0	20	0	0
8.2	Sum of weights after adjustment					
	a. Not a duplicate number	22,583	14,081	34,062	46,754	73,431
	b. Duplicate number	0	0	0	0	0
8.3	Sum of weights	22,583	14,081	34,062	46,754	73,431
8.4	Sample size	897	1,076	800	688	742
8.5	Coefficient of variation	33.15	12.23	19.60	27.44	29.28
9.	Section G nonresponse adjustment					
9.1	Sum of weights before adjustment					
	a. Household with child 1st procedure	877	1,425	1,899	2,393	2,970
	b. Household w/o child 1st proc.-Sect. G comp.	14,251	7,280	19,905	30,899	46,018
	c. Household w/o child 1st proc.-Sect. G not comp.	7,455	5,375	12,258	13,462	24,444
9.2	Sum of weights after adjustment					
	a. Household with child 1st procedure	877	1,425	1,899	2,393	2,970
	b. Household w/o child 1st proc.-Sect. G comp.	21,706	12,655	32,163	44,360	70,462
	c. Household w/o child 1st proc.-Sect. G not comp.	0	0	0	0	0
9.3	Sum of weights	22,583	14,081	34,062	46,754	73,431
9.4	Number HHs w/child 1st procedure					
9.5	Number HHs w/o child 1st proc.–Sect. G comp					
9.6	Coefficient of variation	35.69	25.04	23.14	30.19	31.55

Table B-2. Household weighting for the cell sample by area

	All regions	Northern &Sierra counties	Greater Bay area	Sacramento area	San Joaquin Valley
1. Base weight					
1.1 Sample size	43,694	2,688	8,944	2,606	4,610
1.2 Sum of weight	37,052,158	1,234,826	7,841,387	1,567,375	4,217,225
1.3 Coefficient of variation (CV)	25.69	12.22	20.98	13.49	11.27
2. Adjusting for unknown occupied unit					
2.1 Sum of weights before adjustment					
a. Residential – respondent	5,205,408	158,818	1,149,114	236,428	645,180
b. Residential – nonrespondent	11,115,513	295,993	2,370,389	473,598	1,192,729
c. Unknown residential status (NA, NM)	10,282,325	296,896	2,072,035	487,915	947,660
d. Nonresidential	10,448,912	483,119	2,249,849	369,434	1,431,655
2.2 Sum of weights after adjustment	22,628,494	679,446	5,000,986	1,036,186	2,593,125
2.3 Sample size	6,307	348	1,328	405	707
2.4 Coefficient of variation (CV)	86.44	76.90	82.39	81.28	77.42

Table B-2. Household weighting for the cell sample by area (continued)

		Central Coast	Los Angeles	Other Southern California
1.	Base weight			
1.1	Sample size	3,530	10,109	11,207
1.2	Sum of weight	2.467,228	9,498,284	10,225,833
1.3	Coefficient of variation (CV)	32.56	20.26	19.26
2.	Adjusting for unknown occupied unit			
2.1	Sum of weights before adjustment			
	a. Residential – respondent	361,624	1.225,443	1,428,801
	b. Residential – nonrespondent	656.610	3,039,513	3,086,681
	c. Unknown residential status (NA, NM)	657,104	2,802,431	3,018,283
	d. Nonresidential	791,889	2,430,897	2,692,069
2.2	Sum of weights after adjustment	1,468,349	5,680,458	6,169,944
2.3	Sample size	567	1,355	1,597
2.4	Coefficient of variation (CV)	96.39	85.85	81.96

Table B-3. Household weighting for the area sample by Los Angeles Service Planning Area

		All Los Angeles County	Antelope Valley	San Fernando	San Gabriel	Metro
1.	Base weight					
1.1	Sample size	4,289	122	933	733	601
1.2	Sum of weight	3,220,825	94,668	701,063	540,305	454,700
1.3	Coefficient of variation (CV)	8.52	21.93	6.73	5.31	8.16
2.	Adjusting for unknown occupied unit					
2.1	Sum of weights before adjustment					
	a. Occupied – respondent	997,794	30,687	189,261	210,059	122,476
	b. Occupied – nonrespondent	1,919,552	55,601	482,054	293,863	308,117
	c. Unknown occupied status	209,300	4,960	14,723	20,794	7,545
	d. Not occupied	94,180	3,420	15,024	15,588	16,563
2.2	Sum of weights – allocating unknown occupied					
	a. Occupied – respondent	997,794	30,687	189,261	210,059	122,476
	b. Occupied – nonrespondent	1,919,552	55,601	482,054	293,863	308,117
	c. Unknown occupied status	202,331	4,771	14,401	20,170	7,265
2.3	Sum of weights after adjustment					
	a. Occupied – respondent	997,794	30,687	189,261	210,059	122,476
	b. Occupied – nonrespondent	2,121,883	60,372	496,455	314,034	315,382
	c. Unknown occupied status	0	0	0	0	0
2.4	Sum of weights	3,119,676	91,060	685,716	524,093	437,858
2.5	Sample size	3,887	112	894	684	569
2.6	Coefficient of variation	13.83	23.32	6.77	6.20	8.27

Table B-3. Household weighting for the area sample by Los Angeles Service Planning Area (continued)

	All Los Angeles County	Antelope Valley	San Fernando	San Gabriel	Metro
3. Screener nonresponse adjustment					
3.1 Sum of weights before adjustment					
a. Respondents	1,000,916	30,687	190,020	210,795	123,276
b. Nonrespondents	2,118,760	60,372	495,696	313,298	314,582
3.2 Sum of weights after adjustment					
a. Respondents	3,119,676	91,060	685,716	524,093	437,858
b. Nonrespondents	0	0	0	0	0
3.3 Sum of weights	3,119,676	91,060	685,716	524,093	437,858
3.4 Sample size	1,332	41	251	285	163
3.5 Coefficient of variation	24.23	28.07	5.72	5.34	9.43

Table B-3. Household weighting for the area sample by Los Angeles Service Planning Area (continued)

		West	East	South	South Bay
1.	Base weight				
1.1	Sample size	373	373	480	674
1.2	Sum of weight	300,462	272,899	365,305	491,423
1.3	Coefficient of variation (CV)	4.51	5.89	3.19	11.76
2.	Adjusting for unknown occupied unit				
2.1	Sum of weights before adjustment				
	a. Occupied – respondent	76,439	112,249	142,622	114,001
	b. Occupied – nonrespondent	149,286	129,379	205,181	296,071
	c. Unknown occupied status	64,937	16,457	8,451	71,432
	d. Not occupied	9,800	14,815	9,052	9,919
2.2	Sum of weights – allocating unknown occupied				
	a. Occupied – respondent	76,439	112,249	142,622	114,001
	b. Occupied – nonrespondent	149,286	129,379	205,181	296,071
	c. Unknown occupied status	62,235	15,506	8,236	69,745
2.3	Sum of weights after adjustment				
	a. Occupied – respondent	76,439	112,249	142,622	114,001
	b. Occupied – nonrespondent	211,521	144,885	213,417	365,816
	c. Unknown occupied status	0	0	0	0
2.4	Sum of weights	287,960	257,134	356,039	479,818
2.5	Sample size	279	331	457	561
2.6	Coefficient of variation	16.37	8.67	3.86	14.30

Table B-3. Household weighting for the area sample by Los Angeles Service Planning Area (continued)

	West	East	South	South Bay
3. Screener nonresponse adjustment				
3.1 Sum of weights before adjustment				
a. Respondents	76,439	112,249	143,449	114,001
b. Nonrespondents	211,521	144,885	212,590	365,816
3.2 Sum of weights after adjustment				
a. Respondents	287,960	257,134	356,039	479,818
b. Nonrespondents	0	0	0	0
3.3 Sum of weights	287,960	257,134	356,039	479,818
3.4 Sample size	95	155	188	154
3.5 Coefficient of variation	4.47	6.77	3.19	8.90

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples)

	All Strata	Los Angeles	San Diego	Orange	Santa Clara
1. Adult initial weights					
1.1 Number of sampled adults	85,715	20,977	8,567	5,282	2,891
1.2 Sum of weights	19,308,843	5,265,936	1,560,242	1,638,279	922,414
1.3 Coefficient of variation (CV)	74.56	69.12	53.78	62.10	73.93
2. Nonresponse adjustment					
2.1 Number of completed interviews	49,242	11,201	4,899	2,882	1,689
2.2 Sum of weights before adjustment	19,308,843	5,265,936	1,560,242	1,638,279	922,414
a. Eligible respondents	10,025,957	2,508,768	812,886	811,052	508,584
b. Ineligible	320,242	110,352	26,231	32,572	9,884
c. Nonrespondents	8,962,644	2,646,816	721,125	794,654	403,947
2.3 Sum of weights after adjustment	19,308,843	5,265,936	1,560,242	1,638,279	922,414
a. Eligible respondents	18,528,999	4,971,055	1,496,926	1,555,751	900,166
b. Ineligible	779,844	294,882	63,316	82,528	22,249
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	96.56	95.25	76.15	83.53	90.31
2.5 Mean adjustment factor	0.52	0.48	0.52	0.50	0.55
3. Trimming adjustment*					
3.1 Number of trimmed records	78	17	6	4	4
3.2 Sum of weights before trimming adjustment	18,528,999	4,971,055	1,496,926	1,555,751	900,166
3.3 Sum of weights after trimming adjustment	18,494,637	4,963,483	1,495,853	1,522,253	916,518
4. Raking adjustment*					
4.1 Number of completed interviews	49,242	11,204	4,904	2,829	1,736
4.2 Sum of weights after adjustment	26,873,723	7,327,595	2,198,319	2,255,945	1,320,853
4.3 Coefficient of Variation (CV)	131.39	122.07	101.44	122.36	116.23
4.4 Mean adjustment factor	1.45	1.48	1.47	1.48	1.44
4.5 Mean weight	545.75	654.02	448.27	797.44	760.86

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
1. Adult initial weights					
1.1 Number of sampled adults	3,075	3,178	2,578	2,373	1,722
1.2 Sum of weights	960,769	1,008,337	830,584	716,264	549,950
1.3 Coefficient of variation (CV)	46.36	49.70	54.74	51.56	48.80
2. Nonresponse adjustment					
2.1 Number of completed interviews	1,688	1,749	1,587	1,463	1,055
2.2 Sum of weights before adjustment	960,769	1,008,337	830,584	716,264	549,950
a. Eligible respondents	489,327	501,807	461,965	407,209	308,354
b. Ineligible	15,170	13,643	5,386	12,103	7,597
c. Nonrespondents	456,271	492,887	363,233	296,952	233,999
2.3 Sum of weights after adjustment	960,769	1,008,337	830,584	716,264	549,950
a. Eligible respondents	924,659	972,183	819,200	690,306	535,647
b. Ineligible	36,109	36,154	11,384	25,958	14,304
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	69.17	76.42	70.03	67.49	65.70
2.5 Mean adjustment factor	0.51	0.50	0.56	0.57	0.56
3. Trimming adjustment*					
3.1 Number of trimmed records	0	0	0	0	1
3.2 Sum of weights before trimming adjustment	924,659	972,183	819,200	690,306	535,647
3.3 Sum of weights after trimming adjustment	932,343	974,208	776,914	685,499	583,563
4. Raking adjustment*					
4.1 Number of completed interviews	1,701	1,768	1,520	1,461	1,139
4.2 Sum of weights after adjustment	1,382,111	1,403,051	1,133,530	1,002,685	775,334
4.3 Coefficient of Variation (CV)	92.16	122.26	110.28	98.64	107.17
4.4 Mean adjustment factor	1.48	1.44	1.46	1.46	1.33
4.5 Mean weight	812.53	793.58	745.74	686.30	680.71

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Fresno	San Francisco	Ventura	San Mateo	Kern
1. Adult initial weights					
1.1 Number of sampled adults	1,408	1,575	1,209	1,225	1,145
1.2 Sum of weights	432,940	474,860	406,955	418,056	353,912
1.3 Coefficient of variation (CV)	48.98	61.98	54.63	45.31	55.19
2. Nonresponse adjustment					
2.1 Number of completed interviews	796	933	729	733	676
2.2 Sum of weights before adjustment	432,940	474,860	406,955	418,056	353,912
a. Eligible respondents	222,769	255,274	217,745	228,232	186,472
b. Ineligible	8,540	6,613	4,212	5,354	8,174
c. Nonrespondents	201,631	212,972	184,999	184,470	159,266
2.3 Sum of weights after adjustment	432,940	474,860	406,955	418,056	353,912
a. Eligible respondents	413,778	459,915	396,968	406,592	336,259
b. Ineligible	19,163	14,945	9,987	11,464	17,653
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	67.11	92.00	68.81	67.57	69.79
2.5 Mean adjustment factor	0.51	0.54	0.54	0.55	0.53
3. Trimming adjustment*					
3.1 Number of trimmed records	0	2	0	4	0
3.2 Sum of weights before trimming adjustment	413,778	459,915	396,968	406,592	336,259
3.3 Sum of weights after trimming adjustment	414,810	447,855	406,736	398,372	338,392
4. Raking adjustment*					
4.1 Number of completed interviews	806	920	746	720	681
4.2 Sum of weights after adjustment	628,405	673,541	596,476	557,861	532,782
4.3 Coefficient of Variation (CV)	120.10	107.51	118.04	132.96	118.76
4.4 Mean adjustment factor	1.51	1.50	1.47	1.40	1.57
4.5 Mean weight	779.66	732.11	799.57	774.81	782.35

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
1. Adult initial weights					
1.1 Number of sampled adults	1,140	909	1,018	955	1,034
1.2 Sum of weights	324,501	296,742	260,506	187,104	222,764
1.3 Coefficient of variation (CV)	48.86	48.96	49.27	50.73	43.10
2. Nonresponse adjustment					
2.1 Number of completed interviews	604	579	581	594	569
2.2 Sum of weights before adjustment	324,501	296,742	260,506	187,104	222,764
a. Eligible respondents	152,340	175,491	135,266	108,148	115,982
b. Ineligible	6,785	4,719	3,018	3,116	3,985
c. Nonrespondents	165,376	116,532	122,222	75,841	102,797
2.3 Sum of weights after adjustment	324,501	296,742	260,506	187,104	222,764
a. Eligible respondents	306,672	288,495	253,296	179,670	213,240
b. Ineligible	17,829	8,247	7,210	7,434	9,524
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	75.72	64.76	71.21	66.27	56.72
2.5 Mean adjustment factor	0.47	0.59	0.52	0.58	0.52
3. Trimming adjustment*					
3.1 Number of trimmed records	1	0	0	1	1
3.2 Sum of weights before trimming adjustment	306,672	288,495	253,296	179,670	213,240
3.3 Sum of weights after trimming adjustment	303,682	286,780	242,640	179,750	206,126
4. Raking adjustment*					
4.1 Number of completed interviews	603	590	564	593	553
4.2 Sum of weights after adjustment	442,628	354,995	347,832	300,359	297,633
4.3 Coefficient of Variation (CV)	118.05	111.81	107.32	103.58	98.71
4.4 Mean adjustment factor	1.46	1.24	1.43	1.67	1.44
4.5 Mean weight	734.04	601.69	616.72	506.51	538.21

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Tulare	Santa Cruz	Marin	San Luis Obispo	Placer	Merced
1. Adult initial weights						
1.1 Number of sampled adults	1,085	934	883	831	949	1,073
1.2 Sum of weights	182,253	157,482	159,782	119,206	208,145	101,847
1.3 Coefficient of variation (CV)	43.14	52.46	50.51	52.09	50.93	45.65
2. Nonresponse adjustment						
2.1 Number of completed interviews	583	583	574	578	572	577
2.2 Sum of weights before adjustment	182,253	157,482	159,782	119,206	208,145	101,847
a. Eligible respondents	92,758	91,781	98,360	77,225	114,288	50,810
b. Ineligible	2,975	2,527	1,387	1,032	2,488	1,522
c. Nonrespondents	86,519	63,174	60,035	40,949	91,369	49,515
2.3 Sum of weights after adjustment	182,253	157,482	159,782	119,206	208,145	101,847
a. Eligible respondents	176,336	152,371	157,087	117,328	202,073	98,258
b. Ineligible	5,917	5,112	2,696	1,878	6,072	3,589
c. Nonrespondents	0	0	0	0	0	0
2.4 Coefficient of Variation (CV)	58.65	64.16	64.81	65.67	63.75	65.16
2.5 Mean adjustment factor	0.51	0.58	0.62	0.65	0.55	0.50
3. Trimming adjustment*						
3.1 Number of trimmed records	0	1	0	1	0	2
3.2 Sum of weights before trimming adjustment	176,336	152,371	157,087	117,328	202,073	98,258
3.3 Sum of weights after trimming adjustment	177,569	147,102	158,313	117,889	196,762	107,298
4. Raking adjustment*						
4.1 Number of completed interviews	584	569	575	580	572	595
4.2 Sum of weights after adjustment	287,804	197,369	189,302	192,384	238,226	171,317
4.3 Coefficient of Variation (CV)	92.15	109.70	106.91	102.63	103.21	117.54
4.4 Mean adjustment factor	1.62	1.34	1.20	1.63	1.21	1.60
4.5 Mean weight	492.81	346.87	329.22	331.70	416.48	287.93

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Butte	Shasta	Yolo	El Dorado	Imperial
1. Adult initial weights					
1.1 Number of sampled adults	883	879	880	944	1,118
1.2 Sum of weights	120,909	101,304	100,060	105,026	73,914
1.3 Coefficient of variation (CV)	49.45	49.47	55.92	49.02	44.57
2. Nonresponse adjustment					
2.1 Number of completed interviews	595	575	588	580	581
2.2 Sum of weights before adjustment	120,909	101,304	100,060	105,026	73,914
a. Eligible respondents	78,123	63,012	60,727	59,618	34,918
b. Ineligible	1,658	1,270	876	1,711	1,184
c. Nonrespondents	41,128	37,022	38,457	43,697	37,812
2.3 Sum of weights after adjustment	120,909	101,304	100,060	105,026	73,914
a. Eligible respondents	117,885	98,785	98,424	100,914	70,814
b. Ineligible	3,024	2,519	1,636	4,112	3,101
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	60.95	65.97	69.14	61.70	70.26
2.5 Mean adjustment factor	0.65	0.62	0.61	0.57	0.47
3. Trimming adjustment*					
3.1 Number of trimmed records	2	0	5	3	0
3.2 Sum of weights before trimming adjustment	117,885	98,785	98,424	100,914	70,814
3.3 Sum of weights after trimming adjustment	118,025	102,571	97,809	104,696	69,595
4. Raking adjustment*					
4.1 Number of completed interviews	606	603	595	590	574
4.2 Sum of weights after adjustment	164,068	135,751	138,952	137,505	118,811
4.3 Coefficient of Variation (CV)	105.17	93.44	117.35	86.78	97.04
4.4 Mean adjustment factor	1.39	1.32	1.42	1.31	1.71
4.5 Mean weight	270.74	225.13	233.53	233.06	206.99

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Napa	Kings	Madera	Monterey	Humboldt
1. Adult initial weights					
1.1 Number of sampled adults	972	1,043	1,053	1,028	897
1.2 Sum of weights	71,635	61,292	68,639	200,948	70,960
1.3 Coefficient of variation (CV)	48.80	47.60	44.45	48.69	49.88
2. Nonresponse adjustment					
2.1 Number of completed interviews	575	585	569	574	602
2.2 Sum of weights before adjustment	71,635	61,292	68,639	200,948	70,960
a. Eligible respondents	39,078	31,403	34,759	103,090	45,023
b. Ineligible	1,212	761	1,414	3,382	1,241
c. Nonrespondents	31,345	29,128	32,466	94,477	24,696
2.3 Sum of weights after adjustment	71,635	61,292	68,639	200,948	70,960
a. Eligible respondents	68,951	59,472	65,715	193,272	68,504
b. Ineligible	2,684	1,820	2,924	7,676	2,456
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	67.72	66.81	59.20	74.52	65.91
2.5 Mean adjustment factor	0.55	0.51	0.51	0.51	0.63
3. Trimming adjustment*					
3.1 Number of trimmed records	2	0	1	8	3
3.2 Sum of weights before trimming adjustment	68,951	59,472	65,715	193,272	68,504
3.3 Sum of weights after trimming adjustment	76,397	59,516	65,250	196,357	70,035
4. Raking adjustment*					
4.1 Number of completed interviews	588	585	559	617	617
4.2 Sum of weights after adjustment	96,318	92,156	98,001	291,004	99,954
4.3 Coefficient of Variation (CV)	120.16	103.34	97.74	118.29	106.20
4.4 Mean adjustment factor	1.26	1.55	1.50	1.48	1.43
4.5 Mean weight	163.81	157.53	175.32	471.64	162.00

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Nevada	Mendocino	Sutter	Yuba	Lake
1. Adult initial weights					
1.1 Number of sampled adults	890	919	954	1,036	897
1.2 Sum of weights	70,155	57,769	44,693	38,692	42,231
1.3 Coefficient of variation (CV)	46.40	51.42	49.80	43.27	53.73
2. Nonresponse adjustment					
2.1 Number of completed interviews	582	614	576	582	572
2.2 Sum of weights before adjustment	70,155	57,769	44,693	38,692	42,231
a. Eligible respondents	43,052	35,809	24,574	20,567	25,029
b. Ineligible	369	645	1,232	563	495
c. Nonrespondents	26,735	21,315	18,887	17,561	16,707
2.3 Sum of weights after adjustment	70,155	57,769	44,693	38,692	42,231
a. Eligible respondents	69,438	56,533	42,242	37,414	41,107
b. Ineligible	718	1,236	2,451	1,278	1,124
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	57.98	64.53	62.05	58.80	67.31
2.5 Mean adjustment factor	0.61	0.62	0.55	0.53	0.59
3. Trimming adjustment*					
3.1 Number of trimmed records	0	2	1	0	0
3.2 Sum of weights before trimming adjustment	69,438	56,533	42,242	37,414	41,107
3.3 Sum of weights after trimming adjustment	69,963	54,507	41,908	35,279	40,328
4. Raking adjustment*					
4.1 Number of completed interviews	575	592	573	544	558
4.2 Sum of weights after adjustment	79,825	67,908	63,951	47,927	50,335
4.3 Coefficient of Variation (CV)	81.40	94.61	100.48	92.59	92.88
4.4 Mean adjustment factor	1.14	1.25	1.53	1.36	1.25
4.5 Mean weight	138.83	114.71	111.61	88.10	90.21

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-4. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	San Benito	Colusa, Glenn, Tehama	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1. Adult initial weights				
1.1 Number of sampled adults	1,076	800	688	742
1.2 Sum of weights	30,902	67,884	85,516	136,482
1.3 Coefficient of variation (CV)	43.12	50.33	47.81	49.75
2. Nonresponse adjustment				
2.1 Number of completed interviews	574	483	476	485
2.2 Sum of weights before adjustment	30,902	67,884	85,516	136,482
a. Eligible respondents	15,809	38,333	55,904	84,233
b. Ineligible	206	324	1,107	1,209
c. Nonrespondents	14,887	29,227	28,505	51,040
2.3 Sum of weights after adjustment	30,902	67,884	85,516	136,482
a. Eligible respondents	30,476	67,125	83,462	134,239
b. Ineligible	426	759	2,053	2,243
c. Nonrespondents	0	0	0	0
2.4 Coefficient of Variation (CV)	64.59	70.51	56.74	58.43
2.5 Mean adjustment factor	0.51	0.56	0.65	0.62
3. Trimming adjustment*				
3.1 Number of trimmed records	0	2	0	4
3.2 Sum of weights before trimming adjustment	30,476	67,125	83,462	134,239
3.3 Sum of weights after trimming adjustment	29,254	64,115	83,461	134,860
4. Raking adjustment*				
4.1 Number of completed interviews	536	460	474	483
4.2 Sum of weights after adjustment	40,508	81,577	112,793	148,040
4.3 Coefficient of Variation (CV)	119.09	108.45	101.49	90.55
4.4 Mean adjustment factor	1.38	1.27	1.35	1.10
4.5 Mean weight	75.57	177.34	237.96	306.50

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples)

	All Strata	Los Angeles	San Diego	Orange	Santa Clara
1. Child initial weights					
1.1 Number of sampled children	13,402	3,117	1,397	834	535
1.2 Sum of weights	4,248,067	1,135,308	352,635	383,159	196,551
1.3 Coefficient of Variation (CV)	107.95	98.95	82.73	129.75	90.62
2. Nonresponse adjustment					
2.1 Number of completed child interviews	11,392	2,593	1,197	705	466
2.2 Sum of weights before adjustment	4,248,067	1,135,308	352,635	383,159	196,551
a. Eligible respondents	3,685,731	972,682	307,241	321,786	172,876
b. Ineligible	45,064	19,137	1,253	2,117	761
c. Nonrespondents	517,272	143,489	44,140	59,256	22,914
2.3 Sum of weights after adjustment	4,248,067	1,135,308	352,635	383,159	196,551
a. Eligible respondents	4,196,795	1,113,250	351,212	380,663	195,712
b. Ineligible	51,272	22,058	1,422	2,496	838
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	105.44	99.24	81.35	116.51	87.92
2.5 Mean adjustment factor	1.15	1.17	1.15	1.19	1.14
3. Trimming adjustment*					
3.1 Number of trimmed records	60	8	7	3	4
3.2 Sum of weights before trimming adjustment	4,196,795	1,113,250	351,212	380,663	195,712
3.3 Sum of weights after trimming adjustment	4,133,215	1,100,439	343,429	364,370	197,970
4. Raking adjustment*					
4.1 Number of completed child interviews	9,913	2,177	1,016	612	411
4.2 Sum of weights after adjustment	6,401,053	1,791,543	522,087	518,430	307,562
4.3 Coefficient of Variation (CV)	111.61	107.23	80.47	100.63	108.39
4.4 Mean adjustment factor	1.55	1.63	1.52	1.42	1.55
4.5 Mean weight	645.72	822.94	513.86	847.11	748.33

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
1. Child initial weights					
1.1 Number of sampled children	592	499	422	331	284
1.2 Sum of weights	273,663	244,406	160,694	139,302	115,153
1.3 Coefficient of Variation (CV)	88.82	89.65	68.28	78.17	71.05
2. Nonresponse adjustment					
2.1 Number of completed child interviews	495	419	364	284	248
2.2 Sum of weights before adjustment	273,663	244,406	160,694	139,302	115,153
a. Eligible respondents	241,448	212,992	144,407	124,104	100,267
b. Ineligible	2,349	7,785	748	1,420	1,626
c. Nonrespondents	29,866	23,628	15,539	13,778	13,261
2.3 Sum of weights after adjustment	273,663	244,406	160,694	139,302	115,153
a. Eligible respondents	271,061	235,507	159,878	137,791	113,384
b. Ineligible	2,602	8,899	817	1,511	1,769
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	89.06	83.83	68.08	76.88	71.03
2.5 Mean adjustment factor	1.13	1.15	1.11	1.12	1.15
3. Trimming adjustment*					
3.1 Number of trimmed records	0	1	1	1	3
3.2 Sum of weights before trimming adjustment	271,061	235,507	159,878	137,791	113,384
3.3 Sum of weights after trimming adjustment	268,929	232,761	148,175	133,071	121,750
4. Raking adjustment*					
4.1 Number of completed child interviews	423	364	301	250	246
4.2 Sum of weights after adjustment	374,551	362,591	241,314	243,401	162,566
4.3 Coefficient of Variation (CV)	83.75	86.18	80.52	79.81	87.08
4.4 Mean adjustment factor	1.39	1.56	1.63	1.83	1.34
4.5 Mean weight	885.46	996.13	801.71	973.60	660.84

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	Fresno	San Francisco	Ventura	San Mateo	Kern
1. Child initial weights					
1.1 Number of sampled children	243	174	205	183	201
1.2 Sum of weights	108,611	67,794	90,395	75,414	107,855
1.3 Coefficient of Variation (CV)	68.31	75.35	77.17	62.24	91.44
2. Nonresponse adjustment					
2.1 Number of completed child interviews	199	148	187	158	168
2.2 Sum of weights before adjustment	108,611	67,794	90,395	75,414	107,855
a. Eligible respondents	92,409	57,353	81,658	64,732	90,490
b. Ineligible	568	396	947	0	874
c. Nonrespondents	15,634	10,046	7,790	10,682	16,491
2.3 Sum of weights after adjustment	108,611	67,794	90,395	75,414	107,855
a. Eligible respondents	107,942	67,325	89,384	75,414	106,867
b. Ineligible	669	469	1,012	0	989
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	65.24	70.94	77.68	64.79	90.80
2.5 Mean adjustment factor	1.18	1.18	1.11	1.17	1.19
3. Trimming adjustment*					
3.1 Number of trimmed records	0	0	2	3	2
3.2 Sum of weights before trimming adjustment	107,942	67,325	89,384	75,414	106,867
3.3 Sum of weights after trimming adjustment	107,757	63,105	89,726	73,799	103,753
4. Raking adjustment*					
4.1 Number of completed child interviews	175	122	169	138	148
4.2 Sum of weights after adjustment	174,844	86,901	138,348	111,058	159,283
4.3 Coefficient of Variation (CV)	80.87	74.89	94.92	74.63	88.13
4.4 Mean adjustment factor	1.62	1.38	1.54	1.50	1.54
4.5 Mean weight	999.11	712.31	818.63	804.77	1076.24

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
1. Child initial weights					
1.1 Number of sampled children	181	136	184	150	171
1.2 Sum of weights	81,742	52,957	56,981	42,619	50,833
1.3 Coefficient of Variation (CV)	85.06	75.96	72.15	67.95	83.94
2. Nonresponse adjustment					
2.1 Number of completed child interviews	153	115	163	134	144
2.2 Sum of weights before adjustment	81,742	52,957	56,981	42,619	50,833
a. Eligible respondents	72,256	45,126	51,938	37,492	44,868
b. Ineligible	1,172	227	177	130	195
c. Nonrespondents	8,314	7,604	4,866	4,997	5,771
2.3 Sum of weights after adjustment	81,742	52,957	56,981	42,619	50,833
a. Eligible respondents	80,437	52,669	56,787	42,469	50,613
b. Ineligible	1,306	288	194	150	219
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	85.80	67.85	71.88	63.85	83.66
2.5 Mean adjustment factor	1.13	1.17	1.10	1.14	1.13
3. Trimming adjustment*					
3.1 Number of trimmed records	0	3	0	0	2
3.2 Sum of weights before trimming adjustment	80,437	52,669	56,787	42,469	50,613
3.3 Sum of weights after trimming adjustment	78,638	51,006	54,142	42,448	47,180
4. Raking adjustment*					
4.1 Number of completed child interviews	134	109	140	121	122
4.2 Sum of weights after adjustment	149,417	75,084	108,949	68,044	70,103
4.3 Coefficient of Variation (CV)	99.93	75.43	82.91	91.83	103.27
4.4 Mean adjustment factor	1.90	1.47	2.01	1.60	1.49
4.5 Mean weight	1115.05	688.85	778.20	562.35	574.62

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

		Tulare	Santa Cruz	Marin	San Luis Obispo	Placer	Merced
1.	Child initial weights						
1.1	Number of sampled children	221	150	110	85	138	189
1.2	Sum of weights	52,513	32,698	29,337	15,890	39,322	28,210
1.3	Coefficient of Variation (CV)	76.79	73.17	133.49	70.45	67.59	89.55
2.	Nonresponse adjustment						
2.1	Number of completed child interviews	183	129	96	77	120	153
2.2	Sum of weights before adjustment	52,513	32,698	29,337	15,890	39,322	28,210
	a. Eligible respondents	46,215	28,657	26,274	14,159	35,014	22,876
	b. Ineligible	232	650	674	0	191	0
	c. Nonrespondents	6,067	3,392	2,389	1,731	4,117	5,334
2.3	Sum of weights after adjustment	52,513	32,698	29,337	15,890	39,322	28,210
	a. Eligible respondents	52,252	31,935	28,613	15,890	39,105	28,210
	b. Ineligible	261	764	725	0	217	0
	c. Nonrespondents	0	0	0	0	0	0
2.4	Coefficient of Variation (CV)	76.18	71.94	138.85	72.64	65.96	76.37
2.5	Mean adjustment factor	1.14	1.14	1.12	1.12	1.12	1.23
3.	Trimming adjustment*						
3.1	Number of trimmed records	2	3	2	0	0	0
3.2	Sum of weights before trimming adjustment	52,252	31,935	28,613	15,890	39,105	28,210
3.3	Sum of weights after trimming adjustment	51,268	30,569	28,444	16,269	40,139	32,372
4.	Raking adjustment*						
4.1	Number of completed child interviews	163	112	86	72	118	143
4.2	Sum of weights after adjustment	88,629	39,133	32,146	32,332	51,296	50,347
4.3	Coefficient of Variation (CV)	72.83	100.98	65.41	88.25	86.19	86.47
4.4	Mean adjustment factor	1.73	1.28	1.13	1.99	1.28	1.56
4.5	Mean weight	543.74	349.40	373.79	449.06	434.71	352.08

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

		Butte	Shasta	Yolo	El Dorado	Imperial
1.	Child initial weights					
1.1	Number of sampled children	120	126	140	137	216
1.2	Sum of weights	20,165	19,969	20,190	18,836	20,366
1.3	Coefficient of Variation (CV)	60.18	84.73	76.36	70.74	97.22
2.	Nonresponse adjustment					
2.1	Number of completed child interviews	109	109	123	124	181
2.2	Sum of weights before adjustment	20,165	19,969	20,190	18,836	20,366
	a. Eligible respondents	18,943	17,047	17,953	16,756	17,875
	b. Ineligible	0	73	0	266	0
	c. Nonrespondents	1,222	2,850	2,237	1,815	2,491
2.3	Sum of weights after adjustment	20,165	19,969	20,190	18,836	20,366
	a. Eligible respondents	20,165	19,881	20,190	18,543	20,366
	b. Ineligible	0	88	0	294	0
	c. Nonrespondents	0	0	0	0	0
2.4	Coefficient of Variation (CV)	59.52	77.72	77.25	72.83	99.45
2.5	Mean adjustment factor	1.06	1.17	1.12	1.12	1.14
3.	Trimming adjustment*					
3.1	Number of trimmed records	3	1	2	1	0
3.2	Sum of weights before trimming adjustment	20,165	19,881	20,190	18,543	20,366
3.3	Sum of weights after trimming adjustment	19,854	19,651	20,596	18,366	20,366
4.	Raking adjustment*					
4.1	Number of completed child interviews	98	96	113	110	163
4.2	Sum of weights after adjustment	28,249	24,943	29,660	22,980	28,573
4.3	Coefficient of Variation (CV)	53.72	73.16	82.63	67.01	82.13
4.4	Mean adjustment factor	1.42	1.27	1.44	1.25	1.40
4.5	Mean weight	288.26	259.82	262.48	208.90	175.29

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	Napa	Kings	Madera	Monterey	Humboldt
1. Child initial weights					
1.1 Number of sampled children	121	206	189	177	121
1.2 Sum of weights	12,021	18,746	19,519	49,694	11,483
1.3 Coefficient of Variation (CV)	57.88	107.71	103.25	72.21	64.34
2. Nonresponse adjustment					
2.1 Number of completed child interviews	103	173	156	142	108
2.2 Sum of weights before adjustment	12,021	18,746	19,519	49,694	11,483
a. Eligible respondents	10,617	15,262	17,375	41,398	10,613
b. Ineligible	67	47	146	0	43
c. Nonrespondents	1,338	3,437	1,999	8,296	828
2.3 Sum of weights after adjustment	12,021	18,746	19,519	49,694	11,483
a. Eligible respondents	11,947	18,680	19,345	49,694	11,438
b. Ineligible	74	66	174	0	45
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	57.02	71.42	100.76	72.12	63.75
2.5 Mean adjustment factor	1.13	1.23	1.12	1.20	1.08
3. Trimming adjustment*					
3.1 Number of trimmed records	0	0	1	2	0
3.2 Sum of weights before trimming adjustment	11,947	18,680	19,345	49,694	11,438
3.3 Sum of weights after trimming adjustment	14,276	18,854	18,943	48,189	12,201
4. Raking adjustment*					
4.1 Number of completed child interviews	92	150	147	131	106
4.2 Sum of weights after adjustment	21,872	27,071	24,876	83,008	17,672
4.3 Coefficient of Variation (CV)	90.24	77.70	86.23	110.11	73.25
4.4 Mean adjustment factor	1.53	1.44	1.31	1.72	1.45
4.5 Mean weight	237.73	180.47	169.23	633.65	166.72

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	Nevada	Mendocino	Sutter	Yuba	Lake
1. Child initial weights					
1.1 Number of sampled children	105	122	152	187	105
1.2 Sum of weights	11,035	10,503	11,275	10,547	6,594
1.3 Coefficient of Variation (CV)	69.34	69.52	88.14	92.40	78.52
2. Nonresponse adjustment					
2.1 Number of completed child interviews	93	102	129	169	94
2.2 Sum of weights before adjustment	11,035	10,503	11,275	10,547	6,594
a. Eligible respondents	10,149	9,086	10,112	9,673	5,945
b. Ineligible	0	0	0	185	0
c. Nonrespondents	886	1,416	1,163	690	649
2.3 Sum of weights after adjustment	11,035	10,503	11,275	10,547	6,594
a. Eligible respondents	11,035	10,503	11,275	10,354	6,594
b. Ineligible	0	0	0	194	0
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	68.18	68.65	86.93	91.68	76.00
2.5 Mean adjustment factor	1.09	1.16	1.12	1.09	1.11
3. Trimming adjustment*					
3.1 Number of trimmed records	0	0	1	0	0
3.2 Sum of weights before trimming adjustment	11,035	10,503	11,275	10,354	6,594
3.3 Sum of weights after trimming adjustment	10,433	10,286	10,516	10,233	6,523
4. Raking adjustment*					
4.1 Number of completed child interviews	79	85	104	147	81
4.2 Sum of weights after adjustment	10,297	12,238	18,407	14,687	7,735
4.3 Coefficient of Variation (CV)	72.40	77.46	76.21	101.90	85.57
4.4 Mean adjustment factor	0.99	1.19	1.75	1.44	1.19
4.5 Mean weight	130.35	143.98	176.99	99.91	95.49

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-5. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	San Benito	Colusa, Glenn, Tehama	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1. Child initial weights				
1.1 Number of sampled children	208	105	71	62
1.2 Sum of weights	8,406	12,871	15,401	16,402
1.3 Coefficient of Variation (CV)	96.96	73.65	79.77	79.09
2. Nonresponse adjustment				
2.1 Number of completed child interviews	165	95	64	55
2.2 Sum of weights before adjustment	8,406	12,871	15,401	16,402
a. Eligible respondents	6,774	11,992	14,172	14,670
b. Ineligible	25	70	376	139
c. Nonrespondents	1,607	808	853	1,593
2.3 Sum of weights after adjustment	8,406	12,871	15,401	16,402
a. Eligible respondents	8,372	12,795	15,003	16,248
b. Ineligible	34	75	398	154
c. Nonrespondents	0	0	0	0
2.4 Coefficient of Variation (CV)	78.23	73.77	80.83	72.93
2.5 Mean adjustment factor	1.24	1.07	1.09	1.12
3. Trimming adjustment*				
3.1 Number of trimmed records	1	0	1	1
3.2 Sum of weights before trimming adjustment	8,372	12,795	15,003	16,248
3.3 Sum of weights after trimming adjustment	8,207	12,576	16,153	15,485
4. Raking adjustment*				
4.1 Number of completed child interviews	141	85	63	50
4.2 Sum of weights after adjustment	10,927	18,777	19,059	20,065
4.3 Coefficient of Variation (CV)	93.48	71.10	84.35	69.15
4.4 Mean adjustment factor	1.33	1.49	1.18	1.30
4.5 Mean weight	77.49	220.90	302.52	401.30

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples)

	All Strata	Los Angeles	San Diego	Orange	Santa Clara
1. Adolescent initial weights					
1.1 Number of sampled adolescents	8,308	1,984	816	495	273
1.2 Sum of weights	2,496,165	678,380	194,451	196,044	112,373
1.3 Coefficient of Variation (CV)	93.62	89.74	80.62	80.05	76.50
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	6,990	1,625	670	418	231
2.2 Sum of weights before adjustment	2,496,165	678,380	194,451	196,044	112,373
a. Eligible respondents	2,196,433	588,416	167,333	174,667	95,929
b. Ineligible	35,166	10,045	2,095	2,830	5,560
c. Nonrespondents	264,565	79,918	25,022	18,548	10,884
2.3 Sum of weights after adjustment	2,496,165	678,380	194,451	196,044	112,373
a. Eligible respondents	2,456,896	667,027	192,042	192,900	106,205
b. Ineligible	39,269	11,353	2,409	3,144	6,168
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	91.60	87.55	79.94	77.35	74.18
2.5 Mean adjustment factor	1.12	1.13	1.15	1.10	1.11
3. Trimming adjustment*					
3.1 Number of trimmed records	55	9	9	1	0
3.2 Sum of weights before trimming adjustment	2,456,896	667,027	192,042	192,900	106,205
3.3 Sum of weights after trimming adjustment	2,372,133	642,987	179,253	185,639	105,851
4. Raking adjustment*					
4.1 Number of completed adolescent interviews	3,638	803	317	196	120
4.2 Sum of weights after adjustment	3,510,705	1,023,656	278,955	273,643	146,551
4.3 Coefficient of Variation (CV)	105.53	102.27	67.89	80.58	88.81
4.4 Mean adjustment factor	1.48	1.59	1.56	1.47	1.38
4.5 Mean weight	965.01	1274.79	879.98	1396.14	1221.26

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples) (continued)

	San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
1. Adolescent initial weights					
1.1 Number of sampled adolescents	364	351	255	218	166
1.2 Sum of weights	149,470	162,941	109,928	86,985	63,645
1.3 Coefficient of Variation (CV)	65.66	81.93	72.33	67.01	73.54
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	302	290	225	195	141
2.2 Sum of weights before adjustment	149,470	162,941	109,928	86,985	63,645
a. Eligible respondents	129,347	145,071	100,929	79,942	56,433
b. Ineligible	549	1,053	151	694	520
c. Nonrespondents	19,575	16,818	8,847	6,349	6,692
2.3 Sum of weights after adjustment	149,470	162,941	109,928	86,985	63,645
a. Eligible respondents	148,851	161,735	109,764	86,237	63,064
b. Ineligible	619	1,207	164	748	581
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	62.55	79.80	71.23	66.59	73.98
2.5 Mean adjustment factor	1.15	1.11	1.09	1.08	1.12
3. Trimming adjustment*					
3.1 Number of trimmed records	1	2	0	0	1
3.2 Sum of weights before trimming adjustment	148,851	161,735	109,764	86,237	63,064
3.3 Sum of weights after trimming adjustment	149,696	156,385	104,844	84,841	64,963
4. Raking adjustment*					
4.1 Number of completed adolescent interviews	158	145	118	110	84
4.2 Sum of weights after adjustment	220,870	219,641	121,990	133,426	92,393
4.3 Coefficient of Variation (CV)	84.38	79.76	99.31	88.53	78.75
4.4 Mean adjustment factor	1.48	1.40	1.16	1.57	1.42
4.5 Mean weight	1397.91	1514.76	1033.81	1212.96	1099.91

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples) (continued)

	Fresno	San Francisco	Ventura	San Mateo	Kern
1. Adolescent initial weights					
1.1 Number of sampled adolescents	155	84	130	115	134
1.2 Sum of weights	65,729	31,084	55,166	45,719	61,427
1.3 Coefficient of Variation (CV)	67.00	64.60	71.37	66.09	76.87
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	127	70	117	109	109
2.2 Sum of weights before adjustment	65,729	31,084	55,166	45,719	61,427
a. Eligible respondents	57,812	25,954	50,398	43,020	51,605
b. Ineligible	0	529	495	1,148	3,058
c. Nonrespondents	7,917	4,601	4,274	1,551	6,765
2.3 Sum of weights after adjustment	66,460	30,353	55,166	45,719	61,427
a. Eligible respondents	66,460	29,765	54,627	44,530	57,991
b. Ineligible	0	589	540	1,189	3,436
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	63.77	55.88	72.15	65.28	74.62
2.5 Mean adjustment factor	1.15	1.15	1.08	1.04	1.12
3. Trimming adjustment*					
3.1 Number of trimmed records	3	2	3	1	0
3.2 Sum of weights before trimming adjustment	66,460	29,765	54,627	44,530	57,991
3.3 Sum of weights after trimming adjustment	66,868	22,519	55,108	37,671	55,466
4. Raking adjustment*					
4.1 Number of completed adolescent interviews	64	30	68	56	61
4.2 Sum of weights after adjustment	94,761	28,555	75,900	53,208	83,329
4.3 Coefficient of Variation (CV)	72.99	110.33	60.38	69.93	82.96
4.4 Mean adjustment factor	1.42	1.27	1.38	1.41	1.50
4.5 Mean weight	1480.64	951.84	1116.17	950.15	1366.04

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples) (continued)

	San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
1. Adolescent initial weights					
1.1 Number of sampled adolescents	99	85	129	92	104
1.2 Sum of weights	43,222	32,534	44,917	25,159	29,246
1.3 Coefficient of Variation (CV)	94.82	51.49	66.75	71.04	63.11
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	80	71	114	77	90
2.2 Sum of weights before adjustment	43,222	32,534	44,917	25,159	29,246
a. Eligible respondents	36,654	28,512	41,878	22,217	26,394
b. Ineligible	383	759	201	540	0
c. Nonrespondents	6,184	3,264	2,838	2,402	2,853
2.3 Sum of weights after adjustment	43,222	32,534	44,917	25,159	29,246
a. Eligible respondents	42,774	31,691	44,703	24,562	29,246
b. Ineligible	447	843	215	596	0
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	73.26	50.82	64.71	71.68	61.87
2.5 Mean adjustment factor	1.17	1.11	1.07	1.11	1.11
3. Trimming adjustment*					
3.1 Number of trimmed records	0	1	0	1	2
3.2 Sum of weights before trimming adjustment	42,774	31,691	44,703	24,562	29,246
3.3 Sum of weights after trimming adjustment	41,757	29,418	42,408	24,003	27,219
4. Raking adjustment*					
4.1 Number of completed adolescent interviews	40	34	61	42	44
4.2 Sum of weights after adjustment	70,113	40,042	55,323	36,566	38,377
4.3 Coefficient of Variation (CV)	69.53	64.25	79.88	81.71	72.43
4.4 Mean adjustment factor	1.68	1.36	1.30	1.52	1.41
4.5 Mean weight	1752.83	1177.70	906.94	870.62	872.20

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples) (continued)

	Tulare	Santa Cruz	Marin	San Luis Obispo	Placer	Merced
1. Adolescent initial weights						
1.1 Number of sampled adolescents	126	89	66	72	92	117
1.2 Sum of weights	27,801	17,553	15,118	13,936	27,169	16,744
1.3 Coefficient of Variation (CV)	52.28	58.32	91.56	70.73	61.98	104.32
2. Nonresponse adjustment						
2.1 Number of completed adolescent interviews	105	73	59	64	77	89
2.2 Sum of weights before adjustment	27,801	17,553	15,118	13,936	27,169	16,744
a. Eligible respondents	24,420	14,819	14,115	12,333	24,015	12,744
b. Ineligible	0	233	0	727	0	182
c. Nonrespondents	3,381	2,501	1,003	876	3,154	3,817
2.3 Sum of weights after adjustment	27,801	17,553	15,118	13,936	27,169	16,744
a. Eligible respondents	27,801	17,281	15,118	13,160	27,169	16,508
b. Ineligible	0	272	0	776	0	236
c. Nonrespondents	0	0	0	0	0	0
2.4 Coefficient of Variation (CV)	51.58	58.33	91.39	72.80	59.87	83.08
2.5 Mean adjustment factor	1.14	1.17	1.07	1.07	1.13	1.30
3. Trimming adjustment*						
3.1 Number of trimmed records	1	0	2	2	3	1
3.2 Sum of weights before trimming adjustment	27,801	17,281	15,118	13,160	27,169	16,508
3.3 Sum of weights after trimming adjustment	27,429	16,322	14,482	11,459	25,341	17,418
4. Raking adjustment*						
4.1 Number of completed adolescent interviews	47	43	36	39	46	55
4.2 Sum of weights after adjustment	44,797	18,449	21,514	21,096	31,091	27,670
4.3 Coefficient of Variation (CV)	58.19	89.54	61.82	58.17	65.09	95.92
4.4 Mean adjustment factor	1.63	1.13	1.49	1.84	1.23	1.59
4.5 Mean weight	953.12	429.05	597.62	540.91	675.89	503.08

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples) (continued)

	Butte	Shasta	Yolo	El Dorado	Imperial
1. Adolescent initial weights					
1.1 Number of sampled adolescents	74	92	98	95	129
1.2 Sum of weights	12,343	16,463	13,785	12,127	11,207
1.3 Coefficient of Variation (CV)	65.30	82.99	64.53	74.84	77.74
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	65	83	91	89	97
2.2 Sum of weights before adjustment	12,343	16,463	13,785	12,127	11,207
a. Eligible respondents	11,179	15,269	13,149	11,689	9,345
b. Ineligible	382	109	0	155	390
c. Nonrespondents	782	1,085	637	282	1,472
2.3 Sum of weights after adjustment	12,343	16,463	13,785	12,127	11,207
a. Eligible respondents	11,935	16,346	13,785	11,968	10,758
b. Ineligible	408	117	0	159	449
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	64.00	79.53	63.57	73.86	70.56
2.5 Mean adjustment factor	1.07	1.07	1.05	1.02	1.15
3. Trimming adjustment*					
3.1 Number of trimmed records	0	0	1	1	0
3.2 Sum of weights before trimming adjustment	11,935	16,346	13,785	11,968	10,758
3.3 Sum of weights after trimming adjustment	11,504	16,598	13,479	12,936	10,463
4. Raking adjustment*					
4.1 Number of completed adolescent interviews	39	51	53	56	58
4.2 Sum of weights after adjustment	19,035	16,784	16,609	16,940	17,946
4.3 Coefficient of Variation (CV)	71.04	83.82	81.51	52.28	63.02
4.4 Mean adjustment factor	1.65	1.01	1.23	1.31	1.72
4.5 Mean weight	488.08	329.09	313.38	302.49	309.42

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples) (continued)

	Napa	Kings	Madera	Monterey	Humboldt
1. Adolescent initial weights					
1.1 Number of sampled adolescents	75	127	106	91	68
1.2 Sum of weights	7,295	10,107	9,521	26,399	6,326
1.3 Coefficient of Variation (CV)	51.49	58.57	69.62	71.71	63.83
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	63	106	84	76	62
2.2 Sum of weights before adjustment	7,295	10,107	9,521	26,399	6,326
a. Eligible respondents	6,598	8,699	8,072	23,074	5,687
b. Ineligible	66	193	412	911	0
c. Nonrespondents	631	1,215	1,038	2,414	640
2.3 Sum of weights after adjustment	7,295	10,107	9,521	26,399	6,326
a. Eligible respondents	7,222	9,887	9,059	25,397	6,326
b. Ineligible	72	220	462	1,002	0
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	48.39	56.03	63.79	71.18	58.36
2.5 Mean adjustment factor	1.09	1.14	1.12	1.10	1.11
3. Trimming adjustment*					
3.1 Number of trimmed records	0	1	1	1	0
3.2 Sum of weights before trimming adjustment	7,222	9,887	9,059	25,397	6,326
3.3 Sum of weights after trimming adjustment	8,322	9,366	8,504	24,690	6,500
4. Raking adjustment*					
4.1 Number of completed adolescent interviews	41	46	49	44	46
4.2 Sum of weights after adjustment	11,554	15,942	16,650	37,249	9,930
4.3 Coefficient of Variation (CV)	61.35	73.39	84.96	97.36	64.36
4.4 Mean adjustment factor	1.39	1.70	1.96	1.51	1.53
4.5 Mean weight	281.81	346.57	339.80	846.56	215.86

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples) (continued)

	Nevada	Mendocino	Sutter	Yuba	Lake
1. Adolescent initial weights					
1.1 Number of sampled adolescents	73	75	110	100	81
1.2 Sum of weights	7,501	6,182	6,845	4,872	4,706
1.3 Coefficient of Variation (CV)	61.29	61.15	55.46	78.31	65.48
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	66	69	97	81	71
2.2 Sum of weights before adjustment	7,501	6,182	6,845	4,872	4,706
a. Eligible respondents	7,154	5,715	6,014	4,138	4,315
b. Ineligible	0	89	270	0	36
c. Nonrespondents	346	378	560	733	355
2.3 Sum of weights after adjustment	7,501	6,182	6,845	4,872	4,706
a. Eligible respondents	7,501	6,087	6,550	4,872	4,667
b. Ineligible	0	95	295	0	38
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	58.27	62.96	51.49	67.24	65.29
2.5 Mean adjustment factor	1.05	1.07	1.09	1.18	1.08
3. Trimming adjustment*					
3.1 Number of trimmed records	0	0	0	0	1
3.2 Sum of weights before trimming adjustment	7,501	6,087	6,550	4,872	4,667
3.3 Sum of weights after trimming adjustment	7,157	6,158	6,349	4,764	4,451
4. Raking adjustment*					
4.1 Number of completed adolescent interviews	33	38	53	35	38
4.2 Sum of weights after adjustment	8,666	7,971	9,625	6,213	5,016
4.3 Coefficient of Variation (CV)	65.41	68.52	72.04	75.69	75.77
4.4 Mean adjustment factor	1.21	1.29	1.52	1.30	1.13
4.5 Mean weight	262.62	209.77	181.61	177.53	131.99

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-6. Extended interview weighting for adolescent interview by stratum (combined RDD and surname list samples) (continued)

	San Benito	Colusa, Glenn, Tehama	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1. Adolescent initial weights				
1.1 Number of sampled adolescents	133	68	55	47
1.2 Sum of weights	4,815	8,793	9,201	10,936
1.3 Coefficient of Variation (CV)	64.28	63.16	70.09	54.25
2. Nonresponse adjustment				
2.1 Number of completed adolescent interviews	109	60	49	44
2.2 Sum of weights before adjustment	4,815	8,793	9,201	10,936
a. Eligible respondents	4,193	8,231	8,528	10,429
b. Ineligible	26	0	236	140
c. Nonrespondents	595	563	438	367
2.3 Sum of weights after adjustment	4,815	8,793	9,128	11,009
a. Eligible respondents	4,785	8,793	8,883	10,863
b. Ineligible	30	0	245	146
c. Nonrespondents	0	0	0	0
2.4 Coefficient of Variation (CV)	62.45	60.78	69.35	53.83
2.5 Mean adjustment factor	1.14	1.07	1.04	1.04
3. Trimming adjustment*				
3.1 Number of trimmed records	0	0	0	3
3.2 Sum of weights before trimming adjustment	4,785	8,793	8,883	10,863
3.3 Sum of weights after trimming adjustment	4,606	8,459	9,839	8,644
4. Raking adjustment*				
4.1 Number of completed adolescent interviews	58	32	32	19
4.2 Sum of weights after adjustment	6,090	9,852	12,447	14,271
4.3 Coefficient of Variation (CV)	78.47	73.58	67.42	44.52
4.4 Mean adjustment factor	1.32	1.16	1.27	1.65
4.5 Mean weight	105.01	307.87	388.96	751.11

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-7. Extended interview weighting cell only adults in the combined landline, list and cell phone sample by region

	All regions	Northern &Sierra counties	Greater Bay area	Sacramento area	San Joaquin Valley
1. Adult initial weights					
1.1 Number of sampled adults	1,518	87	335	82	181
1.2 Sum of weights	3,221,200	94,544	705,806	106,273	402,881
1.3 Coefficient of variation (CV)	48.22	49.34	39.91	34.54	33.36
2. Nonresponse adjustment					
2.1 Number of completed interviews	825	53	180	50	93
2.2 Sum of weights before adjustment	3,221,200	94,544	705,806	106,273	402,881
a. Eligible respondents	1,653,395	56,772	349,201	66,681	193,689
b. Ineligible	40,627	1,988	13,231	0	6,041
c. Nonrespondents	1,527,178	35,784	343,373	39,592	203,151
2.3 Sum of weights after adjustment	3,221,200	94,603	705,910	106,120	402,980
a. Eligible respondents	3,141,487	91,402	680,111	106,120	390,788
b. Ineligible	79,713	3,201	25,799	0	12,192
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	49.28	59.63	35.26	22.98	26.46
2.5 Mean adjustment factor	1.95	1.67	2.02	1.59	2.08
3. Trimming adjustment*					
3.1 Number of trimmed records	31	9	1	6	4
3.2 Sum of weights before trimming adjustment	3,141,487	91,402	680,111	106,120	390,788
3.3 Sum of weights after trimming adjustment	2,865,110	73,726	577,964	157,455	288,034
4. Raking adjustment*					
4.1 Number of completed interviews	825	53	180	50	93
4.2 Sum of weights after adjustment	3,535,832	94,198	699,547	213,343	348,067
4.3 Coefficient of Variation (CV)	115.69	57.28	58.74	60.47	38.46
4.4 Mean adjustment factor	1.23	1.28	1.21	1.35	1.21
4.5 Mean weight	4285.86	1777.32	3886.37	4266.86	3742.65

Table B-7. Extended interview weighting cell only adults in the combined landline, list and cell phone sample by region (continued)

	Central Coast	Los Angeles	Other Southern California
1. Adult initial weights			
1.1 Number of sampled adults	124	344	365
1.2 Sum of weights	203,187	872,914	835,596
1.3 Coefficient of variation (CV)	50.80	49.97	41.30
2. Nonresponse adjustment			
2.1 Number of completed interviews	72	167	210
2.2 Sum of weights before adjustment	203,187	872,914	835,596
a. Eligible respondents	110,898	411,971	464,183
b. Ineligible	1,875	15,137	2,354
c. Nonrespondents	90,414	445,806	369,058
2.3 Sum of weights after adjustment	203,403	872,533	835,651
a. Eligible respondents	200,047	841,587	831,433
b. Ineligible	3,356	30,947	4,219
c. Nonrespondents	0	0	0
2.4 Coefficient of Variation (CV)	43.35	49.58	40.71
2.5 Mean adjustment factor	1.83	2.12	1.80
3. Trimming adjustment*			
3.1 Number of trimmed records	1	8	2
3.2 Sum of weights before trimming adjustment	200,047	841,587	831,433
3.3 Sum of weights after trimming adjustment	211,438	735,932	820,561
4. Raking adjustment*			
4.1 Number of completed interviews	72	167	210
4.2 Sum of weights after adjustment	249,191	888,891	1,042,594
4.3 Coefficient of Variation (CV)	57.67	50.44	54.19
4.4 Mean adjustment factor	1.18	1.21	1.27
4.5 Mean weight	3460.99	5322.70	4964.73

Table B-8. Area sample adult extended interview weighting by Los Angeles County Service Planning Area

	All Los Angeles County	Antelope Valley	San Fernando	San Gabriel	Metro
1. Adult initial weights					
1.1 Number of sampled adults	1,332	41	251	285	163
1.2 Sum of weights	6,583,810	184,973	1,419,601	1,169,256	934,021
1.3 Coefficient of variation (CV)	53.26	56.13	47.44	49.33	50.24
2. Nonresponse adjustment					
2.1 Number of completed interviews	981	33	186	219	118
2.2 Sum of weights before adjustment	6,583,810	184,973	1,419,601	1,169,256	934,021
a. Eligible respondents	4,542,890	142,264	977,941	857,512	631,335
b. Ineligible	3,314	0	0	0	0
c. Nonrespondents	2,037,606	42,710	441,660	311,744	302,686
2.3 Sum of weights after adjustment	6,583,810	184,973	1,419,601	1,169,256	934,021
a. Eligible respondents	6,579,242	184,973	1,419,601	1,169,256	934,021
b. Ineligible	4,568	0	0	0	0
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of Variation (CV)	53.25	62.78	43.06	48.84	53.39
2.5 Mean adjustment factor	1.45	1.30	1.45	1.36	1.48
3. Trimming adjustment*					
3.1 Number of trimmed records	17	2	3	2	4
3.2 Sum of weights before trimming adjustment	6,579,242	6,579,242	6,579,242	6,579,242	6,579,242
3.3 Sum of weights after trimming adjustment	6,504,260	6,504,260	6,504,260	6,504,260	6,504,260
4. Raking adjustment*					
4.1 Number of completed interviews	981	33	186	219	118
4.2 Sum of weights after adjustment	7,327,595	198,334	1,577,479	1,336,560	910,164
4.3 Coefficient of Variation (CV)	77.83	74.46	69.17	83.04	79.65
4.4 Mean adjustment factor	1.13	1.12	1.12	1.15	1.00
4.5 Mean weight	7469.52	6010.13	8481.07	6103.01	7713.26

* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Table B-8. Area sample adult extended interview weighting by Los Angeles County Service Planning Area (continued)

	West	East	South	South Bay
1. Adult initial weights				
1.1 Number of sampled adults	95	155	188	154
1.2 Sum of weights	547,876	542,967	800,946	984,170
1.3 Coefficient of variation (CV)	42.08	51.83	46.64	51.07
2. Nonresponse adjustment				
2.1 Number of completed interviews	73	110	126	116
2.2 Sum of weights before adjustment	547,876	542,967	800,946	984,170
a. Eligible respondents	396,448	339,058	487,654	710,678
b. Ineligible	0	0	0	3,314
c. Nonrespondents	151,428	203,910	313,291	270,177
2.3 Sum of weights after adjustment	547,876	542,967	800,946	984,170
a. Eligible respondents	547,876	542,967	800,946	979,602
b. Ineligible	0	0	0	4,568
c. Nonrespondents	0	0	0	0
2.4 Coefficient of Variation (CV)	44.43	49.91	46.90	54.32
2.5 Mean adjustment factor	1.38	1.60	1.64	1.38
3. Trimming adjustment*				
3.1 Number of trimmed records	2	0	3	1
3.2 Sum of weights before trimming adjustment	6,579,242	6,579,242	6,579,242	6,579,242
3.3 Sum of weights after trimming adjustment	6,504,260	6,504,260	6,504,260	6,504,260
4. Raking adjustment*				
4.1 Number of completed interviews	73	110	126	116
4.2 Sum of weights after adjustment	536,865	658,957	948,731	1,160,505
4.3 Coefficient of Variation (CV)	53.29	80.24	69.92	79.89
4.4 Mean adjustment factor	1.00	1.21	1.20	1.19
4.5 Mean weight	7354.31	5990.52	7529.61	10004.35

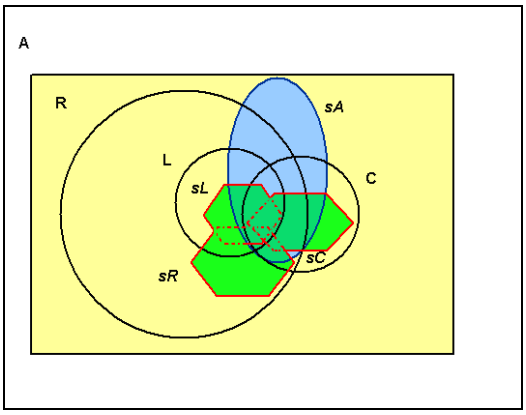
* Counts of completed interviews and sums of weights in sections 3 and 4 are based on self-reported strata rather than sampling strata.

Appendix C

C-1. COMBINING SAMPLES

In this Appendix, we describe how the samples were combined to create the sets of combined weights for CHIS 2007. Before explaining the approach for combining the samples, we examine the relationship between the different frames and samples.

Consider the different samples as illustrated in Figure C-1, which shows the relationship for the Los Angeles County stratum.



Note: The figure is not drawn to scale. The sizes of the samples relative to the frames are smaller than shown in the figure.

Figure C-1. Landline, list, cell phone and area frames, and samples in CHIS 2007

Let A be all eligible households in Los Angeles County (represented by the large yellow rectangles in the diagram). Let R be the landline frame (the large circle in the diagram)) that includes all the households that have a landline telephone[15] and L be the surname list frame (small circle enclosed within R). Note that by definition, R is included within A and that L is included within R (i.e., $L \subset R$). Let C be the frame of the households with cell phones, and thus includes all households with no landline but with one or more cell phones, (i.e., $C \cap \overline{R}$), and households with both types of telephone service (i.e., $C \cap R$). Notice that the cell frame, C, is not encompassed by R, but crosses both R and A. Let s_A, s_R, s_L, and s_C be the area, landline, surname list, and cell phone samples respectively.

As shown in the figure, the landline, cell phone, and area frames overlap. Since the landline, cell-phone, and area samples are probability samples, we used a multiple frame estimation approach to combine and create weights for these samples. This approach follows the ideas of Hartley (1962) (although we do not use the estimator proposed in this paper), and it is different from the approach used to combine the landline/surname

[15] The geographic supplemental sample for San Diego County is considered part of the landline sample.

sample or the combined landline/list/cell phone sample. We used this method because the method used for the other combined samples requires knowing the multiple probabilities of selection of all units in the sample, which could not be determined in this situation (we would need to be able to identify the area sample cluster that includes the telephone from the landline and cell phone samples).

Instead, we created separate final weights for the area sample (s_A) and the combined landline-cell phone samples ($s_R \cup s_c$) by applying the appropriate weighting adjustments (See *CHIS 2007 Methodology Report 1 -Weighting and Variance Estimation).* Since these landline, cell phone, and area samples were drawn separately, we can produce estimates for different domains of adults. There are three domains of interest in the overlapping frames. The first domain is all adults in households with landline telephone service (i.e., $r \in R$), the second domain is all adults in cell-only households (i.e., $c \in C \cap \overline{R}$), and the third domain is all adults in households without any type of telephone service (i.e., $a \in A \cap \left(\overline{R \cup C}\right)$). The landline/cell sample and area sample overlap in the first two domains, while the last domain contains only records from the area sample. Because of these overlaps, the expression of the weights is different for the last domain.

Let Y be a characteristic for all adults in Los Angeles County (e.g., the number of adults with health insurance). Let $\hat{Y}_{comb}$ be the estimate of Y computed using the combined sample, let $\hat{Y}_R$ be the estimate the estimate of Y computed using the records in combined landline-list-cell phone sample and let $\hat{Y}^A$ be the estimate of Y computed using the area sample. An unbiased estimate of Y is computed as

$$\hat{Y}_{comb} = \begin{cases} \lambda \hat{Y}^R + (1-\lambda)\hat{Y}^A & \text{If respondent has telephone service of any kind} \\ \hat{Y}^A & \text{If respondent does not have telephone service} \end{cases},$$

where λ $(0 \le \lambda \le 1)$ is the composite or weighting factor. In other words, for common domains among the samples, $\hat{Y}_{comb}$ is the weighted average of the estimates $\hat{Y}^R$ and $\hat{Y}^A$. Hartley (1962, 1974) proposed that the value of λ be computed to minimize a statistic such the variance of $\hat{Y}_{comb}$. Because the samples are drawn independently, the estimate of variance of $\hat{Y}_{comb}$ is

$$\hat{V}\left(\hat{Y}_{comb}\right) = \hat{\lambda}^2 \hat{V}\left(\hat{Y}^R\right) + \left(1-\hat{\lambda}\right)^2 \hat{V}\left(\hat{Y}^A\right).$$

The Hartley approach has several disadvantages. First, since the value λ depends on the variable Y, we need to compute a different value of λ, for each variable from the survey. Different values imply different weights, and the estimates from the different sets of weights would not be additive (e.g., the sum of males and females is not the same as the total). Furthermore, we need to compute the estimates from the two

samples before we can produce the composite estimate. These obstacles make the production of estimation for a large number of variables using this approach impractical.

We used an alternative approach where the sampling weights were composited, rather than the estimates. In this approach a single compromise value of λ is attached to the final weights before combining the samples. The composite weights are then used to compute estimates for any variable. For example, the expression for the estimate $\hat{Y}_{comb}$ becomes

$$\hat{Y}_{comb} = \sum_{i \in s_R \cup s_L \cup s_C} \hat{\lambda} w_i^R y_i^R + \sum_{i \in s_A} \left(1 - \hat{\lambda}\right) w_i^A y_i^A \ ,$$

where $\hat{\lambda}$ is the composite factor. Note that the new weights for the combined landline-list-cell phone samples are $\hat{\lambda}\, w_i^R$ and the new weights for the area sample cases (without a landline or a cell phone households) are $\left(1 - \hat{\lambda}\right) w_i^A$.

In this approach, the weights for the combined samples are created in an extra weighting step after the other final weights (for the weights described above and in the report). The advantage of this approach is that it uses a unique value of $\hat{\lambda}$, and once the compositing factor is attached to the weights the computation of estimates is transparent to the user. With the appropriate replicate structure, the computation of the estimates of variance is also transparent to the user. These can be computed in the same way as they are currently produced without any modification of the replication software.

C-2. WEIGHTING PROCESS

Figure 2 shows the steps for creating the weight for the combined sample. The process began with the files for the landline, cell phone and area samples containing the person level weights adjusted for nonresponse at the household and person level①. These files were created in the weighting process as described in Chapters 3 through 6 of *CHIS 2007 Methodology Report 5 -Weighting and Variance Estimation.*

In the next step, the landline/cell phone sample is separated into two files, one with cases in Los Angeles County (LA) and the second with the remainder of California. Since the area sample includes only cases in LA, there is no need to separate this sample②. Next ③, the area sample is raked to control totals defined for LA. After raking the area sample, the weights of these cases sum to the population in LA.

After raking, the area cases were separated into three files ④. The first file included the area records with no telephone service of any type. The second file contained the area records with a landline, and the last file contains the records with only cell phone service. The weights from the last two files (landline and cell-only) were used to derive sample-based control totals for raking the combined landline and cell phone sample in LA ⑤. The sample-based control totals represented the total number of adults in LA that had telephone service (either landline or cell phone). These control totals were derived from the area records with telephone service. They are called sample-based because the totals have sampling error and that error is reflected in the combined landline and cell phone sample replicate weights.

In the next step, the weights of the combined landline and cell phone sample in LA were raked to the sample based control totals derived from the area sample ⑥. After this raking step, the weights of cases from the area and RDD samples that have telephone service sum to the same total.

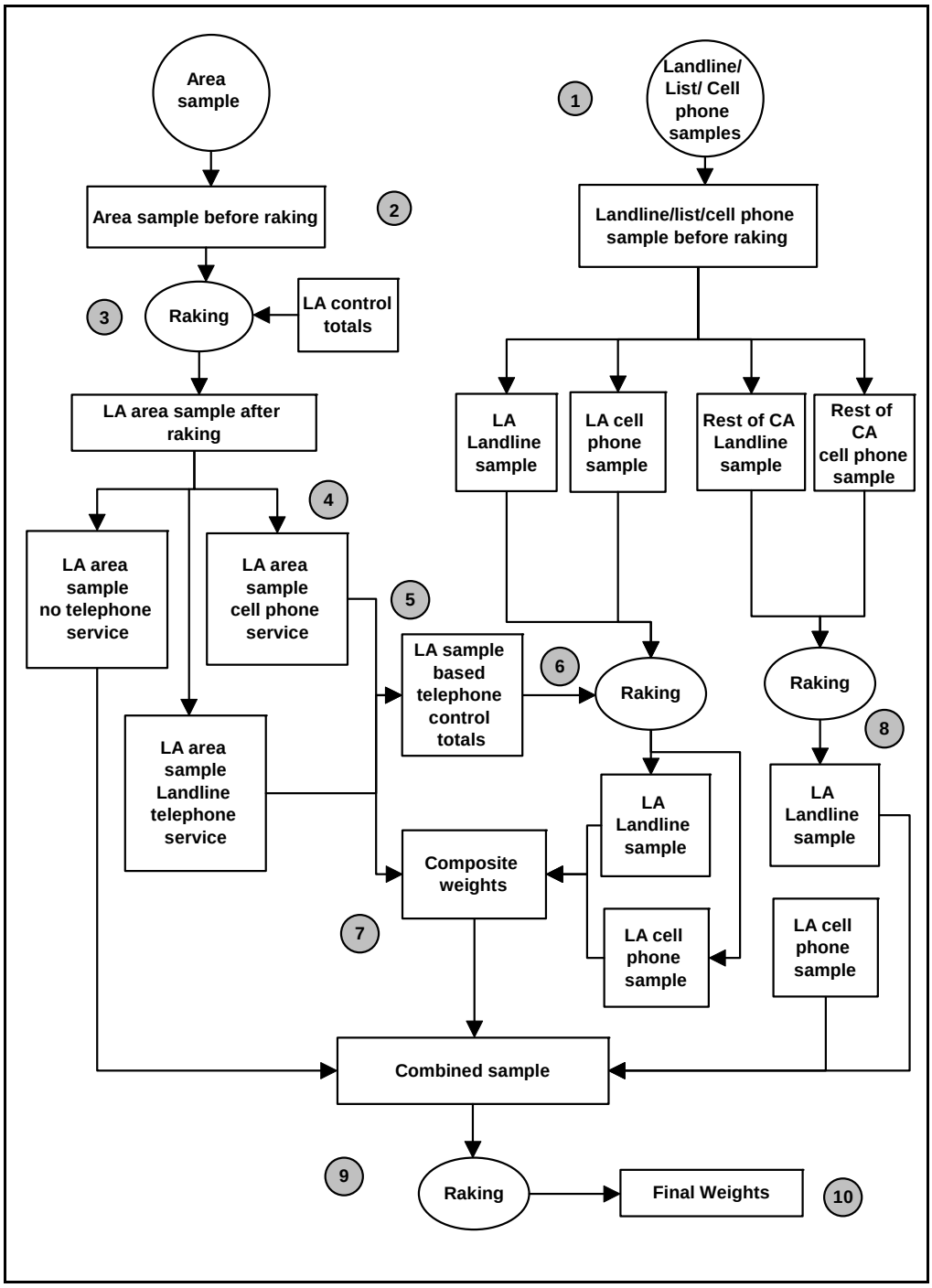


Figure C-2. Weighting the landline, list, cell phone, and area samples

In the following step ⑦, we created the composite factor that combined the weights of the cases from the area and RDD samples. The weights were combined by type of telephone (i.e., landline or cell phone only). Since the samples have been raked, the sum of weights was similar between the samples so the composite factor did not adjust for bias. Table C-1 shows the value of lambda computed minimizing different criteria. We averaged 113 estimates separately by landline/cell phone and area samples to compute values of $\hat{\lambda}$ for the

variance and effective sample size. The different values of $\hat{\lambda}$ were similar, and we used $\hat{\lambda} = 0.90$ as the composite factor in CHIS 2007[16].

Table C-1. Values of lambda

Criterion	$\hat{\lambda}$
Sample size	0.93
Variance	0.92
Effective sample size	0.90

Table C-2 shows the sum of weights before and after the composite weights for the area and combined landline/cell phone sample

Table C-2. Adult sums of weights and percentages by type of telephone service before and after the composite factor.

Before composite weights

	Sum of weights			Percentage		
Telephone service	Landline/cell phone sample	Area	Total	Landline/cell phone sample	Area	Total
Landline only	2,039,378	1,817,366	3,856,744	29%	25%	27%
Cell phone only	932,865	556,383	1,489,248	13%	8%	10%
Both	4,162,271	4,760,764	8,923,035	58%	67%	63%
Total	7,134,514	7,134,513	14,269,027	100%	100%	100%

After composite weights

	Sum of weights			Percentage		
Telephone service	Landline/cell phone sample	Area	Total	Landline/cell phone sample	Area	Total
Landline only	1,794,652	218,084	2,012,736	29%	25%	28%
Cell phone only	820,921	66,766	887,687	13%	8%	12%
Both	3,662,798	571,292	4,234,090	58%	67%	59%
Total	6,278,371	856,142	7,134,513	100%	100%	100%

Weights from cases in the area sample with no telephone service were not raked or combined in this or any previous steps because the landline-cell phone sample did not include cases without telephone service.

After attaching the composite factor ⑧, the weights from the landline-cell phone sample in the remainder of the state are raked to control totals that exclude the population in Los Angeles County. This step

[16] Since both samples add to the total adults in Los Angeles, there is a factor of 0.5 applied to the two samples before applying the composite factor

brought the weights to the same level as the landline-cell phone sample in LA (excluding the cases with no telephone service).

After the third raking, a single file was created by combining the files with the LA composite weights (landline and cell-phone only cases from the area and landline-samples), LA cases with no telephone service, and the files with landline-cell phone cases in the remainder of the state ⑨. In the last step, the weights in the combined file were raked to state control totals ⑩.

Table C-3 in the shows the sum of weights and sample size before and after the adjustments after each step.

Table C-3. Sample sizes, sum of weights, and coefficient of variations by sample type in the combined sample.

Weighting process	Landline/cell	Area	Total
1. Before first raking			
a. Sample size	11,370	981	12,351
b. Sum of weights	5,739,402	6,579,242	12,318,644
c. Coefficient of variation	139	53	-
2. After raking before composite weight			
a. Sum of weights	7,134,514	7,134,513	14,269,027
b. Coefficient of variation	156	78	-
3.After composite weight before final fake			
a. Sum of weights	6,421,062	906,534	7,327,596
b. Coefficient of variation	156	132	155
4. After final fake			
a. Sum of weights	6,420,635	906,961	7,327,595
b. Coefficient of variation	157	131	155

C-3. VARIANCE ESTIMATION

Replicates are used to compute estimates of variances based on resampling methods. There were 80 replicate weights in the CHIS 2007. The replicates were created using the same paired unit jackknife method (JK2[17]) as in the other CHIS samples. The same variance estimation strata for the area and combined landline-cell phone sample were used. Since the final file was created by appending the two samples, the number of replicates stayed the same. Although the samples are independent and the correlation between the estimates from the two samples is zero, some correlation could be induced in the estimates when the two samples are raked to the same control totals. Because of this possibility, we decided to create 160 replicate weights (as in the replicates used to estimate differences between years). Using 160 replicates forces the correlation to be zero. For additional details on the creation of the replicate weights for the landline, cell and area samples, see Chapter 9 in *CHIS 2007 Methodology Report 5-Weighting and Variance Estimation.*

We also created the variable TSVARSTR required to compute estimates of variance using linearization. The variable TSVARSTR (Taylor's series variance stratum) indicates the variance stratum and was created using the replicate variance stratum with values 1 to 80 from the landline/list and cell phone samples. While we believe this is a reasonable approach for linearization variances, linearization programs do not reflect the full range of weighting steps used in the production of the composite estimates.

[17] This method is denoted as JK2 in the software program, WesVar, which was used to compute all the sampling errors in this report.