CALIFORNIA HEALTH INTERVIEW SURVEY

CHIS 2003 METHODOLOGY SERIES

REPORT 5

WEIGHTING AND VARIANCE ESTIMATION

August 2006

This report was prepared for the California Health Interview Survey by Ismael Flores Cervantes, J. Michael Brick, Brandon Barrett, R. Lee Harding, and Michael E. Jones of Westat.

www.chis.ucla.edu

This report describes the weighting and variance estimation methods used in CHIS 2003. 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 2003 Methodology Series: Report 5 – Weighting and Variance Estimation*. Los Angeles, CA: UCLA Center for Health Policy Research, 2005.

The California Health Interview Survey is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. Funding for CHIS 2003 came from multiple sources: the California Department of Health Services, the California Endowment, the National Cancer Institute, the Robert Wood Johnson Foundation, the California Office of the Patient Advocate, the U. S. Centers for Disease Control and Prevention, Kaiser Permanente, the Alameda County Health Care Agency, and L.A. Care Health Plan.

PREFACE

Weighting and Variance Estimation is the fifth in a series of methodological reports describing the 2003 California Health Interview Survey (CHIS 2003). 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 Health Services, and the Public Health Institute. Westat was responsible for the data collection and the preparation of five methodological reports from the 2003 survey. The survey examines public health and health care access issues in California. The CHIS telephone survey is the largest state health survey ever undertaken in the United States. The plan is to monitor the health of Californians and examine changes over time by conducting periodic surveys in the future.

Methodological Reports

The first five methodological reports for CHIS 2003 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 2003. 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

List of Tables

TABLE OF CONTENTS (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Figures

1. CHIS 2003 DESIGN AND METHODOLOGY SUMMARY

1.1 Overview

The California Health Interview Survey (CHIS) is a population-based random-digit dial telephone survey of California's population that is conducted every two years. First conducted in 2001, CHIS is the largest health survey ever conducted in any state and one of the largest health surveys in the nation. CHIS is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health 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 development issues.

The CHIS sample is designed to provide population-based estimates for most California counties, all major ethnic groups, and several ethnic subgroups. The sample is designed to meet and optimize two goals: provide estimates for large- and medium-sized population counties in the state, and for groups of the smallest population counties; and provide statewide estimates for California's overall population, its major race/ethnic groups, as well as for several Asian ethnic groups. The resulting 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 the 2003 California Health Interview Survey (CHIS 2003). CHIS 2001 is described in a series of methodology reports.[1] These reports describe the second CHIS data collection cycle, which was conducted between August 2003 and February 2004.

CHIS data and results are used extensively by many State agencies, local public health agencies and organizations, federal agencies, advocacy and community organizations and agencies, foundations, and researchers. They use these data in their own analyses and publications to assess public health and health care needs, to develop health policies, and to develop and advocate policies to meet those needs.

[1] California Health Interview Survey, CHIS 2001 Methodology Series: 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, Los Angeles, CA: UCLA Center for Health Policy Research, 2002.

1.2 Sample Design Objectives

The CHIS sample is designed to meet two objectives: (1) provide estimates for counties and groupings of counties with populations of 100,000 or more; and (2) provide estimates for California's overall population and its larger race/ethnic groups, as well as for several smaller ethnic groups. To achieve these objectives, CHIS relied on a multi-stage sample design. First, the state was divided into 41 geographic sampling strata, including 33 single-county strata and 8 groups that included the 25 other counties. Second, within each geographic stratum, households were selected through random-digit dial (RDD), and within each household, an adult (age 18 and over) respondent was randomly selected. In addition, in those households with adolescents (ages 12-17) and/or children (under age 12), one adolescent was randomly selected for interview and one child was randomly selected and the most knowledgeable parent of the child interviewed.

Table 1-1 shows the 41 sampling strata (i.e., counties and groups of counties that were identified in the sample design as domains for which separate estimates would be produced). A sufficient amount of sample was allocated to each of these domains to support the first sample design objective. These strata were also used for the CHIS 2001 sample; because of funding limitations, the sample sizes allocated to most strata for CHIS 2003 were smaller than in 2001.

Table 1-1. California county and county group strata used in the CHIS 2003 sample design

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

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

The samples in Los Angeles and Alameda Counties were enhanced with additional funding to allow sub-county geographic estimates, in Los Angeles at the Service Planning Area (SPA) level and in Alameda for the cities of Oakland and Hayward as well as the remainder of the county. These samples were implemented with and incorporated into the original statewide RDD sample.

To accomplish the second objective, larger sample sizes were allocated to the more urban counties where a significant portion of the state's Latino, African American and Asian ethnic populations reside. To increase the precision of the estimates for Koreans and Vietnamese, areas with relatively high concentrations of these groups were sampled at higher rates; these geographic samples were supplemented by phone numbers for group-specific surnames drawn from listed telephone directories to increase the sample size and precision of the estimates for these two groups.

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 research that identified 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 2003 data collection. Westat staff interviewed one randomly selected adult in each sampled household. In those households with children (under age 12) or adolescents (ages 12-17) associated with the sampled adult[2], one child and one adolescent were randomly sampled, so up to three interviews could have been completed in each sampled household. The sampled adult was interviewed, and the parent or guardian most knowledgeable about the health and care of the sampled child was interviewed. The sampled adolescent responded for him or herself, but only after a parent or guardian gave permission for the interview. Table 1-2 shows the number of completed adult, child, and adolescent interviews in CHIS 2003, by the type of sample (RDD or supplemental sample).

[2] Only children for whom the sampled adult was parent or legal guardian were sampled. The CHIS 2003 sample weights account for this sampling procedure.

Table 1-2. Number of completed CHIS 2003 interviews by type of sample, instrument

Type of sample	Adult	Child	Adolescent
Total RDD + supplemental cases	42,044	8,526	4,010
RDD	41,818	8,480	3,996
Supplemental samples:			
Korean	112	24	6
Vietnamese	114	22	8

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

Interviews done in all languages were administered using Westat's computer-assisted telephone interviewing (CATI) system. The average adult interview took 33 minutes to complete. The average child and adolescent interviews took 14 minutes and 21 minutes, respectively. Interviews in the non-English languages generally took longer to complete. Approximately 11 percent of the adult interviews were completed in a language other than English, as were 21 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 Rate

The overall response rate for CHIS 2003 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 the selected person to complete the full 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 72 percent of the sampled telephone numbers. In 2003, the screener completion rate was 55.9 percent[3], and the rate was higher for those households that could be sent the advance letter. The extended interview completion rate was 60.0 percent for the adult survey. Multiplying the screener and extended rates gives an overall response rate of 33.5 percent. Response rates vary by sampling stratum.

[3] In CHIS 2003, households that refused at the screener level were subsampled and only the subsampled households were called again in an attempt to convert them to respondents. The response rates are weighted to account for this subsampling.

Table 1-3. CHIS 2003 survey topic areas by instrument

Health Status	**Adult**	**Teen**	**Child**
General health status, height and weight	✓	✓	✓
Emotional health		✓	
Days missed from school due to health problems		✓	✓
Health Conditions	**Adult**	**Teen**	**Child**
Asthma	✓	✓	✓
Heart disease, high blood pressure, epilepsy	✓		
Diabetes	✓	✓	
Physical disability/need for special equipment	✓	✓	✓
Elder health (stroke, falls, incontinence)	✓		
Parental concerns with child development, attention deficit disorder (ADD)			✓
Health Behaviors	**Adult**	**Teen**	**Child**
Dietary intake		✓	✓
Physical activity and exercise		✓	✓
Walking for transportation and leisure	✓		
File and pneumonia immunization	✓		
Alcohol and tobacco use	✓	✓	
Drug use		✓	
Sexual behavior, STD testing, birth control practices	✓	✓	
Women's Health	**Adult**	**Teen**	**Child**
Pap test screening, mammography screening, self-breast exam	✓		
Emergency contraception, pregnancy status	✓	✓	
Menopause, hormone replacement therapy (HRT)	✓		
Cancer History and Prevention	**Adult**	**Teen**	**Child**
Cancer history of respondent	✓		
Colon cancer screening, prostrate cancer (PSA) test	✓		
Dental Health	**Adult**	**Teen**	**Child**
Last dental visit, could not afford care, missed school/work days	✓	✓	✓
Dental insurance coverage	✓	✓	✓
Injury/Violence	**Adult**	**Teen**	**Child**
Serious injuries (frequency, cause)		✓	✓
Injury prevention behaviors (bike helmets, seatbelts)		✓	✓
Infant-toddler home safety			✓
Interpersonal violence		✓	

Table 1-3. CHIS 2003 survey topic areas by instrument (Continued)

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)	✓	✓	✓
Health care discrimination due to race or ethnic group	✓		
Communication problems with doctor	✓	✓	✓
Ability and parental knowledge of teen contacting a doctor		✓	
Child immunization reminders			✓
Health Insurance	**Adult**	**Teen**	**Child**
Current insurance coverage, spouse's coverage, who pays for it	✓	✓	✓
Health plan enrollment, characteristics and assessment of plan	✓	✓	✓
Whether employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months	✓	✓	✓
Reasons for lack of insurance	✓	✓	✓
Employment	**Adult**	**Teen**	**Child**
Employment status, spouse's employment status	✓		
Work in last week, industry and occupation	✓		
Hours worked at all jobs	✓	✓	
Income	**Adult**	**Teen**	**Child**
Respondent and spouse's earnings last month before taxes	✓		
Household income (annual before taxes)	✓		
Number of persons supported by household income	✓		
Assets	✓		
Public Program Eligibility	**Adult**	**Teen**	**Child**
Household poverty level (100%, 130%, 200%, 300% FPL)	✓		
Program participation (TANF, CalWorks, Public Housing, Food Stamps, SSI, SSDI, WIC)	✓	✓	✓
Assets, alimony/child support/social security/pension	✓		
Reason for Medi-Cal non-participation among potential eligibles	✓	✓	✓
Food Insecurity/Hunger	**Adult**	**Teen**	**Child**
Availability of food in household over past 12 months	✓		
Parental Involvement	**Adult**	**Teen**	**Child**
Parental presence after school, parental knowledge of whereabouts and activities		✓	
Child's activities with family			✓

Table 1-3. CHIS 2003 survey topic areas by instrument (Continued)

Neighborhood and Housing	**Adult**	**Teen**	**Child**
Neighborhood cohesion	✓		
Neighborhood safety	✓	✓	
Neighborhood characteristics for children			✓
Length of time at current address/neighborhood, type of housing	✓		
Home ownership, number of rooms, amount of mortgage/rent	✓		
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)	✓	✓	✓
Health care discrimination due to race or ethnic group	✓		
Communication problems with doctor	✓	✓	✓
Ability and parental knowledge of teen contacting a doctor		✓	
Child immunization reminders			✓
Health Insurance	**Adult**	**Teen**	**Child**
Current insurance coverage, spouse's coverage, who pays for it	✓	✓	✓
Health plan enrollment, characteristics and assessment of plan	✓	✓	✓
Whether employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months	✓	✓	✓
Reasons for lack of insurance	✓	✓	✓
Employment	**Adult**	**Teen**	**Child**
Employment status, spouse's employment status	✓		
Work in last week, industry and occupation	✓		
Hours worked at all jobs	✓	✓	
Income	**Adult**	**Teen**	**Child**
Respondent and spouse's earnings last month before taxes	✓		
Household income (annual before taxes)	✓		
Number of persons supported by household income	✓		
Assets	✓		
Public Program Eligibility	**Adult**	**Teen**	**Child**
Household poverty level (100%, 130%, 200%, 300% FPL)	✓		
Program participation (TANF, CalWorks, Public Housing, Food Stamps, SSI, SSDI, WIC)	✓	✓	✓
Assets, alimony/child support/social security/pension	✓		
Reason for Medi-Cal non-participation among potential eligibles	✓	✓	✓
Food Insecurity/Hunger	**Adult**	**Teen**	**Child**
Availability of food in household over past 12 months	✓		

Table 1-3. CHIS 2003 survey topic areas by instrument (Continued)

Parental Involvement	**Adult**	**Teen**	**Child**
Parental presence after school, parental knowledge of whereabouts and activities		✓	
Child's activities with family			✓
Neighborhood and Housing	**Adult**	**Teen**	**Child**
Neighborhood cohesion	✓		
Neighborhood safety	✓	✓	
Neighborhood characteristics for children			✓
Length of time at current address/neighborhood, type of housing	✓		
Home ownership, number of rooms, amount of mortgage/rent	✓		
Child Care	**Adult**	**Teen**	**Child**
Current child care arrangements			✓
Child care over past 12 months			✓
Reason for lack of childcare			✓
Respondent Characteristics	**Adult**	**Teen**	**Child**
Age, gender, height, weight, education	✓	✓	✓
Race and ethnicity	✓	✓	✓
Marital status	✓		
Sexual orientation	✓		
Citizenship, immigration status, country of birth, English language proficiency	✓	✓	✓

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

The CHIS response rate is comparable to response rates of other scientific telephone surveys in California, such as the California Behavioral Risk Factor Surveillance System (BRFSS) survey. 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.

One way to judge the representativeness of a population survey is to "benchmark" its results against those of other reliable data sources. The CHIS 2001 sample yielded unweighted and weighted population distributions and rates that are comparable to those obtained from other sources. The demographic characteristics of the CHIS 2001 sample (such as race, ethnicity, and income) are very similar to those obtained from 2000 Census data. CHIS 2001 respondents also have health characteristics and behaviors that also are very similar to those found in other reliable surveys, such as the California

BRFSS. An extensive benchmarking project is being undertaken for the 2003 California Health Interview Survey.

Adults who had completed at least 80 percent of the questionnaire (i.e., through Section I on health insurance) after all followup 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 these cases.

Proxy interviews were allowed for frail and ill persons over the age of 65 to avoid biases for health estimates for elderly persons that might otherwise result. Eligible selected persons were recontacted and offered a proxy option. For 171 elderly adults, a proxy interview was completed by either a spouse/partner or adult child. Only a subset of questions identified as appropriate for a proxy respondent were administered. (Note: The questions not administered are identified in their response set as being skipped (denoted by a value of "-2") because a proxy is responding for the selected person.)

1.5 Weighting the Sample

To produce population estimates for the RDD CHIS results, weights are applied to the sample data to compensate for a variety of 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. Sample weighting was carried out in CHIS 2003 to accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons (Note: telephone numbers for which addresses could be found and advance letters mailed were assigned a higher probability of selection than those without addresses);
- 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" or the

inverse of the probability of selection of the telephone number and adjustment factors computed for the following weight adjustments:

- Subsampling for numbers with addresses;
- Multiple chances of being selected in the RDD and supplemental samples;
- Unknown residential status;
- Subsampling screener refusals for conversion attempt;
- Screener interview nonresponse;
- Multiple telephone numbers; and
- Household poststratification.

The resulting poststratified household weight was used to compute a person-level weight. This person-level weight includes weight 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 totals from auxiliary data sources. The procedure requires iteration to make sure all the control totals or dimensions of raking are simultaneously satisfied (within a specified tolerance).

The control totals or raking dimensions used in CHIS 2003 were created primarily from the 2003 California Department of Finance estimates of the numbers of persons by age, race, and sex, and from the 2000 Census of Population counts from the U.S. Census Bureau. The 14 dimensions are combinations of demographic variables (age, sex, race, and ethnicity), geographic variables (county, city, and, in Los Angeles County, Service Planning Area), 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 a telephone number from the survey. One of the limitations of using the 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 9 or more unrelated persons). These persons were excluded from the CHIS sample and, as a result, the number of persons living in group quarters had to be estimated and removed from the control totals prior to raking.

1.6 Imputation Methods

To enhance the utility of the CHIS 2003 data files, missing values were replaced through imputation for nearly every variable. This was a massive task designed to eliminate missing values in all source variables. Westat imputed values for variables used in the weighting process, and the UCLA staff imputed values where missing due to item nonresponse for nearly all other variables.

Two different imputation procedures were used by Westat prior to delivering the data to UCLA to fill in missing responses for items in CHIS 2003 that were essential for weighting the data. The first imputation technique is a completely random selection from the observed distribution of the respondents. This method is used only for a few items when the percentage of the items that are missing is very small. For example, when imputing the missing values for self-reported age which had a very low item non-response rate, the distributions of the responses for age by type of interview (adult, child, or adolescent) were used to randomly assign an age using probabilities associated with these distributions.

The second technique is hot deck imputation without replacement. The hot deck approach is probably the most commonly used method for assigning values for missing responses in large-scale household surveys. 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 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 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 race, ethnicity, home ownership, and education in CHIS 2003.

The UCLA staff imputed missing values through a hierarchical sequential hot deck method with donor replacement. This method rank-orders the control variables from the most essential to the least essential, allowing the control variables to be dropped if the imputation conditions (such as minimal number of donors or no missingness in control variables) are not met in the imputation process. The control variables are dropped one at a time sequentially, starting from the least essential. CHIS incorporated an automated data quality control check both before and after the imputation process.

Imputation flags for CHIS source variables are included in separate data files to identify all imputed values.

1.7 Methodology Report Series

A series of five methodology reports are available with more detail about the methods used in CHIS 2003:

- 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 www.CHIS.ucla.edu or contact CHIS at CHIS@ucla.edu.

2. WEIGHTING ADJUSTMENTS CHIS 2003 SAMPLE WEIGHTS

This chapter introduces the concept of weighting and provides some background on the weights developed for analyzing CHIS 2003 survey data. Weighting is a process that attempts to make the estimates from the survey 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 sample. The process begins with a base weight that is then adjusted to account for additional stages of sampling and nonresponse. The base weight is the inverse of the probability of selection of the sampled unit. During the weighting, additional information from external sources is used to benchmark the weights and achieve consistency between totals from the survey data and the external sources.

Although this chapter deals with the weight adjustments for the CHIS 2003 sample, the chapter begins with the general reasons why fully adjusted weights should be used. It also describes the details, advantages, and disadvantages of weighting. In CHIS 2003, the random digit dial (RDD) sample[4] and surname list samples are combined and weighted together in a single file.

2.1 Weighting Approach

Weights are applied to CHIS 2003 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.

In CHIS 2003 one set of weights was created for each instrument or questionnaire (adult, child, and adolescent) for the combined RDD and Korean and Vietnamese surname list samples. Appendix A shows the CHIS 2003 frame and sample sizes in addition to base weights by sampling strata

[4] The geographic supplemental sample in Antelope Valley and Alameda County (Hayward and Oakland) are considered as part of the RDD sample.

for the RDD and Korean and Vietnamese lists. These weights can be used to produce estimates at the state and stratum (county or group of counties) level. Each final weight is the result of a series of sequential adjustments made to the base weights. The details for the creation of the weights for the combined RDD-list sample are given in Chapters 3, 4, 5, and 6.

As part of the weighting process for the RDD samples, a household weight is created for each household that completed the screener interview. This household weight is the base weight computed as the inverse of the probability of selection of the sample telephone number adjusted for:

- Subsampling for mail status;
- Subsampling for refusal conversion;
- Unknown residential status;
- Supplemental sample race eligibility;
- Screener interview nonresponse;
- Multiple telephone numbers; and
- Household poststratification.

The details of these adjustments are described in Chapter 3.

The poststratified household weight is adjusted to create a person weight for each type of extended interview. The final person weight incorporates the within-household probability of selection of the sampled person and accounts for nonresponse. Each adjustment corresponds to a multiplicative weighting factor applied to the weight. 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 somewhat more complex because of the sampling method used (see *CHIS 2003 Methodology Series: Report 1 - Sample Design*). For these weights, the factors include:

- Section G adult extended interview nonresponse (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; 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 are described in Chapter 8. Chapter 9 describes the methods for variance estimation for CHIS 2003.

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.

2.2 Weight Adjustment Method

In an ideal survey, all the units in the inference population are eligible to be selected into the sample and all those that are selected participate in the survey. In practice, neither of these conditions occurs. Some units are not eligible for the sample (undercoverage) and some of the sampled units do not respond (nonresponse). If undercoverage and nonresponse are not addressed, then estimates from the survey will be biased. In CHIS 2003, the weights of those who respond are adjusted to represent the undercovered persons and nonrespondents. An overview of the approaches used to account for these two sources of missing data begins 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 on 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 2003. 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, we know the county in California associated with each telephone number, 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 degree of bias reduction in the adjustment is larger as either response rates or the survey characteristics are more similar within the cells.

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. Very 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 2003 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 of such importance (as well as the city-level estimates for some cities), 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 uncovered 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 2003 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 poststratification loosely and intend it to include raking, a form of multidimensional poststratification (see Brackstone and Rao, 1979). In CHIS 2003, the control totals are mainly derived from the 2003 California Department of Finance Population Projections (State of California, Department of Finance, 2004), the 2002 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). Totals for the number of households with or without someone under 18 years old by sampling stratum were used as a control in the last

household adjustment to create the household weights. Details of the creation of the control totals at the person level are described in Chapter 7.

The next chapters describe how these approaches were applied in the weighting procedures used for the CHIS 2003 combined RDD and surname list samples.

3. HOUSEHOLD WEIGHTING

The first step in the weighting process for the combined RDD and surname list samples of CHIS 2003 is creating a household weight for each completed screener interview. Although the household weight is not used for analytical purposes because the only data captured at the household level in the screener interview are mainly 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 eight sections, each describing the steps involved in creating the household weights. The first section reviews the creation of base weights as the inverse of the probability of selection of the telephone number. Subsequent sections describe the adjustments made to the base household weights. These adjustments account for subsampling based on mail status, refusal conversion, unknown residential status, screener interview nonresponse, and households with multiple telephone numbers. The final section of the chapter describes how the household weights are poststratified to control totals for the number of households in California.

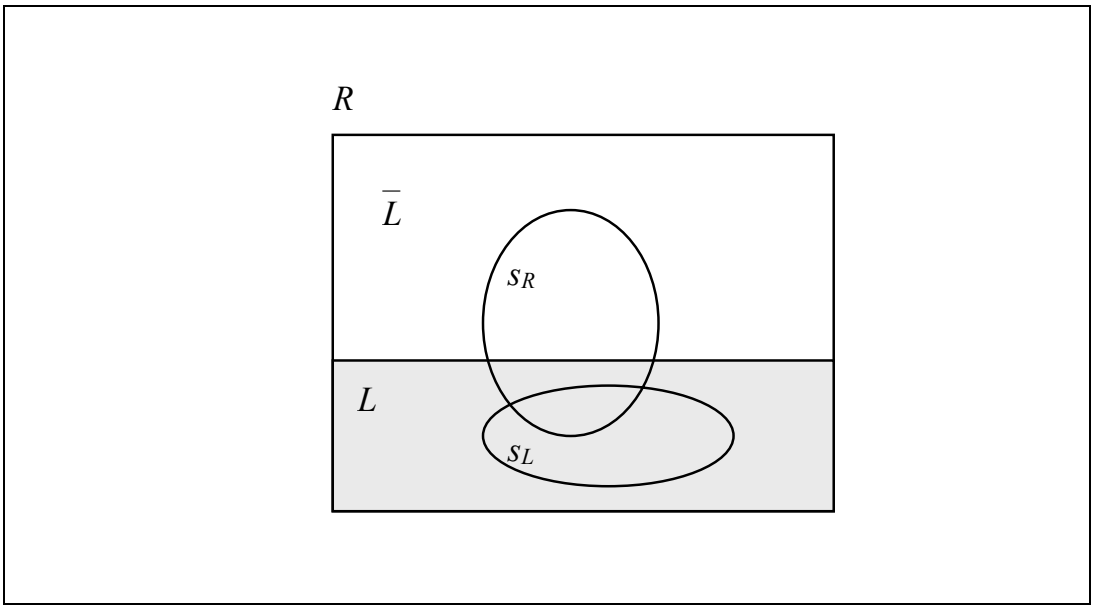
A background in the sampling methods used in CHIS 2003 is an essential ingredient to understanding the weighting procedures. The first report in this series describes the sampling and we assume anyone interested in the weighting procedures is already familiar with the contents of Report 1: Sample Design. We do briefly restate some of the sampling methods in the course of this document, but do not give details.

3.1 Base Weights

Each telephone number in the sample is assigned a base weight. Base weights are computed as the inverse of the probability of selection of the telephone number. In CHIS 2003, telephone numbers were drawn from three frames (RDD frame, Korean and Vietnamese surname lists); therefore, the base weights reflect the multiple probabilities of selection of telephone numbers from these different frames.

Figure 3-1 shows the relationship between the RDD frame and a single list frame (i.e., Korean or Vietnamese list) for a single sampling stratum (i.e., county, city, or group of counties). The figure also shows the relationship of the type of samples drawn from each frame. In order to create the household base weights, we consider all telephone households in California as either being on the

supplemental list only (L) or as only being eligible for sampling from the RDD 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 RDD samples (s_R and s_L) are smaller than shown in the figure.

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

The notation in the figure follows:

R	the RDD frame containing all telephone numbers.
L	the list frame (i.e. surnames, of telephone numbers in Zip Codes, etc).
$\overline{L}$	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.
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) in s_r.
n_L	the sample size (number of telephone numbers) in s_L.

Note that the RDD 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. Note that $n_R = n_{RL} + n_{R\bar{L}}$.

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 base weights can be expressed as:

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

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

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

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

Creating these weights required being able to classify 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 RDD 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 number of telephone numbers in the RDD frame and list frames (N_R and N_L) are computed separately. The RDD sample was drawn using a list-assisted approach from a stratified frame of 100 banks[5] with at least one listed telephone number in the state of California. Using this approach, a

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

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 RDD 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 July 2002 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 RDD sampling strata by linking telephone exchanges to the counties in the same way as for the RDD sample. The list size by stratum (N_{Lh}) is the number of records in the list assigned to stratum h.

As described in the CHIS 2003 Sample Design Report (Report 1), the RDD sample was drawn from strata defined as counties or groups of counties except for Los Angeles, San Diego, Orange, Santa Clara, and Alameda Counties. In Los Angeles County, 12 subsampling strata were created by the combination of areas with high concentration of Koreans and Vietnamese and Special Planning Areas (SPAs). Substrata based on the concentration of Koreans and Vietnamese were created for San Diego, Orange, and Santa Clara Counties. Because of separate sample size goals by city and race in Alameda County, 9 substrata were created based on the concentration of households in the cities of Oakland and Hayward and the concentration of African Americans in Hayward. Appendix A shows the number of telephone numbers in the frame, the number of sample cases, and base weights by frame type (RDD, Korean and Vietnamese lists), sampling stratum, and substratum. Table B-1 in Appendix B (rows 1.1 and 1.2) lists the sample counts and sums of base weights by sampling stratum.

3.2 Subsampling Adjustment

To reduce the number of calls to ineligible telephone numbers in household surveys (nonworking and business numbers) and improve the efficiency of the sample, special procedures were implemented before data collection. In one of these procedures telephone numbers were classified as listed or unlisted by automatically matching the sampled numbers to the White Pages (residential numbers) and Yellow Pages (business numbers). In addition to the listed status, each telephone number was classified by whether a mailing address could be associated with it.[6] We refer to those telephone numbers associated with a mailing address as having a "mailable" address. A second procedure involved

[6] Several companies provide services of this type in which a telephone number is matched to commercially available files of addresses.

a tritone purging method to identify the working status of a telephone number (working or nonworking). Telephone numbers classified as nonresidential or nonworking were not dialed for CHIS 2003.

The efficiency of the sample was further improved by stratifying the telephone numbers by mail status and subsampling the strata at different rates (Brick, Judkins, Montaquila, and Morganstein, 2002). Mailable telephone numbers are more likely to be residential, so the cost of finding a residence is much lower in the substratum of mailable numbers. In addition, households with mailable telephone numbers are more likely to cooperate with most surveys[7] (Brick et al., forthcoming). The stratified samples were subsampled at rates determined using the principles of optimal allocation to balance both data collection costs and the variances of the estimates.

Substrata were created using the original strata and the information on telephone number working status (residential, business, or nonworking) and mail status (the telephone number has a "mailable" address or not) as shown in Table 3-3. The telephone numbers not selected in the subsample were eliminated and never dialed. Nonmailable telephone numbers were subsampled at a rate of 0.75[8] in all strata/substrata except for telephone numbers in Alameda County[9] or from the Korean and Vietnamese supplemental list samples.

Mailable and nonmailable telephone numbers in the nine sampling strata within Alameda County were subsampled at different rates. The subsampling rates in the mailable/nonmailable substrata in the areas that covered Hayward, Oakland, and the remainder of Alameda County were optimized in order to yield the required number of completed interviews[10] in these areas minimizing the design effect. The list samples were not subsampled because all list telephone numbers had a mailing address.

At the end of the data collection, as the sampling goals for some strata were being met, only a random portion of the telephone numbers already drawn were released for dialing. These partial releases were done within strata separately by mailable substratum. These partial releases resulted in achieving different subsampling rates by mailable substratum than was planned. The observed mailable subsampling rates are shown in Table 3-1. Table B-1 in Appendix B (rows 2.7a through 2.7c) shows the observed subsampling rates by sampling stratum.

---

[7] The subsampling increases the percentage of respondents but not the response rates since the response rates are weighted to account for the subsampling. (See Report 4: Response Rates.)

[8] 75 percent of the nonmailable households were kept in the RDD sample.

[9] The sample drawn for Alameda County includes the Oakland and Hayward geographic supplemental samples and the Hayward African American supplemental sample.

[10] The respondent's reported location was used to determine the number of completed cases for the areas in Alameda County.

Table 3-1. Mailable substratum definition*

Mailable substrata		Planned subsampling rate	Observed subsampling rate	Description
M	Mailable	100.0%	91.8%	Telephone numbers with a "mailable" address not classified as non-working or business
NM	Nonmailable	75.0%	60.1%	Telephone numbers without a "mailable" address not classified as non-working or business
NR	Nonresidential	75.0%	75.0%	Telephone numbers identified as business or nonworking

* Calculation of the observed rates excludes telephone numbers in Alameda County and from the Korean and Vietnamese list samples.

[a] Although nonresidential telephone numbers were not dialed, they are kept in the files to facilitate the computation of residency rates during data collection.

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

The mailable subsampling adjusted household weight, $HHA1W_i$, is computed as:

$$HHA1W_i = HHA1F_c * HHBSW_i$$

where $HHA1F_c$ is the mailable subsampling adjustment factor computed as:

$$HHA1F_c = \begin{cases} \dfrac{\sum HHBSW_i\delta_i(c)}{\sum_{i\in INSMP} HHBSW_i\delta_i(c)} & \text{If } i \in INSMP \\ 0 & \text{otherwise} \end{cases},$$

where *INSMP* is the set of telephone numbers retained in the mailable subsample, and $\delta_i(c)$ is 1 if the number is in mailable substratum *c* as defined in Table 3-1 and is zero otherwise. The subsampling weighting adjustment is done within sampling stratum, but for notational convenience a subscript for the stratum is not included in the formulas. Table B-1 of Appendix B (row 2.5) shows the sum of the weights after the subsampling adjustment. The telephone numbers in nonresidential substratum *NR* identified in the tritone purge were dropped from the weighting process at this point because they are not residential numbers.

3.3 Refusal Subsampling Adjustment

After adjusting for mailable subsampling, the weights were adjusted to reflect the refusal conversion subsampling efforts made during data collection. In CHIS 2003, refusal conversion procedures were applied to a random subsample of screener interview refusals. During sample selection, 60 percent of the telephone numbers were subsampled for refusal conversion. Refusal conversion was attempted if the respondent refused to answer the screener interview only for these telephone numbers. The cases subsampled for refusal follow-up were fielded at the beginning of the data collection so that any refusal cases could be worked completely. In CHIS 2003, refusal cases comprised the majority of screener nonresponse and substantial effort was required to gain their cooperation. By subsampling screener interview refusals, resources were shifted from the less productive, labor-intensive task of refusal conversion to the more productive task of completing extended interviews.

In order to adjust the weights for screener interview refusal subsampling, telephone numbers were classified into screener refusal groups using their refusal status (i.e., if the respondent ever refused or not) and the value of the refusal conversion flag as shown in Table 3-2:

Table 3-2. Screener refusal groups

Screener refusal group	Respondent ever refused screener interview?	Refusal subsampling flag	Description
NRef	No	N/A	Households where respondent did not refuse the screener interview (includes complete and noncomplete interviews)
RefC	Yes	Yes	Households where respondent refused to do the screener interview and refusal conversion procedures were used
RefNC	Yes	No	Households where respondent refused to do the screener interview and refusal conversion procedures were not used

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

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

$$HHA2W_i = HHA2F_i * HHA1W_i$$

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

$$HHA2F_c = \begin{cases} \dfrac{\sum_{i \in RefC, RefNC} HHA1W_i \delta_i(c)}{\sum_{i \in RefC} HHA1W_i \delta_i(c)} & \text{If } i \in RefC \\ 0 & \text{If } i \in RefNC \\ 1 & \text{If } i \in NRef \end{cases},$$

where the groups *RefC*, *RefNC*, and *NRef* are defined in Table 3-2, and $\delta_i(c)$ is 1 if the number is in sampling stratum *c* and is zero otherwise. Refusal conversion procedures were only used in cases from the RDD samples; as a result, no adjustment was needed for cases from the list samples. As described in the previous section, towards the end of the field period, not all sampled telephone numbers were released. This factor also had an impact in the refusal subsampling rate because the cases flagged for refusal conversion were fielded at the beginning of data collection. The observed refusal subsampling rate was 65 percent[11] in CHIS 2003. Table B-1 in Appendix B (rows 3.1a through 3.3) shows the sum of the weights before and after the refusal conversion subsampling adjustment.

3.4 Unknown Residential Status Adjustment

At the end of data collection not all telephone numbers can be classified as residential or not despite being dialed many times. Numbers that cannot be classified are considered as unknown residential status. They are telephone numbers that were answered only by answering machines (screener disposition code of *NM*) or were never answered even by a machine (screener disposition of *NA*, ring no answer). Prior to adjusting the RDD weights for screener interview nonresponse, we estimated the number of eligible residential telephone numbers among those numbers with unknown residential status. This estimate was also used in the computation of the response rates described in Report 4: Response Rates.

In CHIS 2003, the estimated proportion of unknown residential telephone numbers considered residential (p_{res}) was computed separately for the RDD and list samples. The value of p_{res} for the RDD sample was computed using a survival method with censored data (Brick, Montaquila, and Scheuren, 2002). Under this model, the "treatment" is the number of calls made to the telephone number until it is resolved as either residential or not. The data are censored because numbers were not called indefinitely. For the RDD sample, the value of p_{res} was computed using the results of a subsample of 3,511 telephone numbers with unknown residential status (numbers with a screener result code of *NA*

[11] Excluding cases in Alameda County and cases from the Korean and Vietnamese samples.

"ring no answer" or *NM* "answer machine") that was re-released for additional contact attempts. The proportion p_{res} was computed within groups defined by urban status,[12] mailable status of the telephone number, and the answering machine status given by the interviewer (no machine, residential, or nonresidential) based on the content of the machine's message.

Table 3-3 shows the values of p_{res} for the RDD sample computed using the survival analysis method. As expected, the estimated proportion of residential households is much lower for answering machines coded as "nonresidential" compared to those coded as "residential." For example, the estimated proportion of residential households in urban strata with mailable addresses and answering machines coded as residential is 87.9 percent, while the estimated proportion of those coded as urban mailable-nonresidential is 28.9 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 nonresidential.

Table 3-3. Estimated residential proportion for the CHIS 2003 RDD sample

Urban status	Mail status	Answering machine code	p_{res}
Urban	mailable	no machine	0.588
Urban	mailable	residential	0.879
Urban	mailable	nonresidential	0.289
Urban	mailable	unknown	0.774
Urban	nonmailable	no machine	0.074
Urban	nonmailable	residential	0.819
Urban	nonmailable	nonresidential	0.114
Urban	nonmailable	unknown	0.572
Rural	mailable	no machine	0.638
Rural	mailable	residential	0.895
Rural	mailable	nonresidential	0.312
Rural	mailable	unknown	0.663
Rural	nonmailable	no machine	0.059
Rural	nonmailable	residential	0.821
Rural	nonmailable	nonresidential	0.082
Rural	nonmailable	unknown	0.540

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

[12] For this purpose, urban strata are defined as those strata with a population of over 500,000 persons. Stratum 1 (Los Angeles) through Stratum 15 (San Joaquin) in Table 3-1 are urban, while the remaining strata are rural.

In the list sample, where no telephone numbers were redialed, the value of p_{res} was computed in accordance with the standards defined by CASRO (Council of American Survey Research Organizations) for RDD surveys (Council of American Survey Research Organizations, 1982). That is, the value of p_{res} was the observed proportion of residential telephone numbers among the list telephone numbers with known residential status. Table 3-4 shows the estimated proportion of households among list cases with unknown residential status in CHIS 2003.

Table 3-4. Estimated residential proportion for the CHIS 2003 list samples

List Sample	p_{res}
Korean	0.754
Vietnamese	0.780

The estimated proportion of residential households among the unknown residential telephone numbers is 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:

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

where the group *RES* denotes telephone numbers identified as residential and *UNK_RES* denotes telephone numbers with unknown residential status. This adjustment is done separately for the RDD (within sampling stratum) and list samples. After adjusting for telephones with unknown residential status, only telephone numbers known to be households have positive weights. Table B-1 in Appendix B (rows 4.1a through 4.4) shows the sum of weights before and after making the adjustment for unknown residential status.

3.5 Supplemental Sample Race Eligibility Adjustment

Unlike the CHIS 2003 RDD sample where every adult in the household was eligible, adults associated with telephone numbers selected for the ethnic supplemental samples (i.e., a subset of RDD cases in areas with high concentration of African-Americans in Hayward and Korean and Vietnamese list samples) were sampled only if the adults considered themselves to be African-American, or of Vietnamese or Korean descent (Question SC6A1H of the screener interview). If there was an adult from one of these groups in the household (an adult eligible by race or ethnicity, referred to here as a "race-eligible adult"), then one adult was sampled among those eligible within the household. If there were no race-eligible adults, the screener interview was terminated and the case was coded as a race-ineligible household. A special nonresponse adjustment was implemented to account for the supplemental cases where the race eligibility of the adults of the households could not be determined. This adjustment is called the supplemental sample race eligibility adjustment and was implemented only for households screened on race or ethnicity (i.e., African-American in Hayward, Korean, or Vietnamese in the Hayward supplement or the list samples). All remaining cases have an adjustment factor of 1.

Before adjusting the weights for race eligibility, the cases were classified in response groups based on race as indicated in Table 3-5.

Table 3-5. Race eligibility response groups

Race response status group		Description
R_E	Race Eligible	Supplemental sample household with at least one race eligible adult
R_IN	Race Ineligible	Supplemental sample household without any race eligible adult
R_UNK	Race Eligibility Unknown	Supplemental sample household where the race eligibility of the adults could not be determined
R_NA	Race not screened	Household not in the supplemental samples

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

$$HHA4W_i = HHA4F_c * HHA3W_i,$$

where $HHA4F_c$ is the race screener nonresponse adjustment factor computed as

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

where the groups *R_E*, *R_IN*, *R_UNK*, and *R_NA* are defined in Table 3-5, and $\delta_i(c)$ is 1 if the number is in race eligibility screener nonresponse adjustment cell *c* and is zero otherwise. The nonresponse adjustment cells corresponded to the sample type (i.e., Korean list, Vietnamese list, African-American in Hayward sample), and remainder of RDD cases. Table B-1 in Appendix B (rows 5.1a through 5.2c) gives the sum of weights before and after the race eligibility screener nonresponse adjustment.

3.6 Screener Nonresponse Adjustment

After adjusting the weight for unknown residential status, only the telephone numbers for residential households had positive weights. Nevertheless, not all of the residential households completed the screener interview. In this step, the household weight is adjusted to account for households that did not complete the screener interview. The screener nonresponse adjusted household weight, $HHA5W_i$, is:

$$HHA5W_i = HHA5F_c * HHA4W_i,$$

where $HHA5F_c$ is the screener nonresponse adjustment factor computed as

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

where the group *SC_R* is the set of screener respondents, *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. Cases from the supplemental samples with adult race ineligibles (group *R_IN* defined in the previous section) were considered as screener nonrespondents (*SC_R*) in this adjustment. Although these cases were

households with race ineligible adults, they still represented households with eligible adults for the regular CHIS extended interview who were screened out.

The nonresponse adjustment cells were created separately for the RDD and list samples within sampling strata using the telephone mailable status (known address, unknown address). These cells have different response rates due to the effect of the pre-notification letter sent to households with a known address and differences associated with the mailable and nonmailable groups. Table B-1 in Appendix B (rows 6.1 through 6.2) gives the sum of weights before and after the screener nonresponse adjustment.

3.7 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 the additional 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. If this is the case, the household weight is adjusted to reflect the increased probability of selection. The multiple telephone adjusted household weight, $HHA6W_i$, is computed as:

$$HHA6W_i = HHA6F_i * HHA5W_i ,$$

where $HHA6F_i$ is the multiple telephone adjustment factor computed as:

$$HHA6F_i = \begin{cases} 0.5 & \text{If the household has more than one residential telephone number} \\ 1 & \text{Otherwise} \end{cases} .$$

In this adjustment for RDD telephone numbers, 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 CHIS 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 and 7.2) identifies the sum of weights before and after the multiple telephone adjustment.

3.8 Household Poststratification

The final step in weighting the screener interviews was to poststratify the household weights to household control totals as was done in CHIS 2001. The poststratification adjustment in CHIS 2001 used cells created for households with and without a person under 18 years old within sampling stratum. The control totals were derived from the Census 2000 Summary File 1 for California (U.S. Census Bureau, 2001). In CHIS 2003, no data source contained the total number of households by the presence of persons under 18 years old for 2003 for each county in California. However, because it is known that response rates differ between households with and without a person under 18 years old and poststratifying household level weights using cells for households with and without children adjusts for this difference in response, the household weights were poststratified to 2001 control totals. Using the 2000 Census data introduced a bias in the sum of household weights, but no household estimates were produced using the CHIS 2003 data. The bias did not have an effect on person-level estimates because the person weights were benchmarked to 2003 population totals. Therefore any person-level estimate is scaled to the appropriate population total in California as of 2003. See more details in Section 7-3.

The household poststratification weight, $HHA7W_i$, is computed as

$$HHA7W_i = HHA7F_k * HHA6W_i,$$

where $HHA7F_k$ is the poststratification factor for cell *k* computed as

$$HHA7F_k = \frac{CNT_k}{\sum_{i \in k} HHA6W_i},$$

and CNT_k is the control total for cell *k* defined by the existence of persons under 18 years old in the household.

The overall poststratification adjustment factors for the state and by sampling stratum are listed in Table B-1 in Appendix B (row 8.4). Because the control totals were for 2001 rather than 2003, the magnitude of this adjustment should not be used as an indirect measure of the undercoverage of the estimate of the total number of households.

4. ADULT WEIGHTING

Adult final weights were created for each adult who completed the adult extended interview.[13] 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. In subsequent steps, the initial adult weight is adjusted for nonresponse and raked to known control totals. To deal with undercoverage of adults who could not be interviewed because they reside in nontelephone households, the raking adjustment included a dimension designed to reduce the bias from this source. Details on creating the adult weights follow.

4.1 Adult Initial Weight

As described in Section 4.2 of *CHIS 2003 Methodology Series: Report 1 - Sample Design*, one adult was sampled from among all adults in a household with equal probability using the Rizzo method[14] (see Rizzo et. al., 2004). The initial adult weight is computed as the product of the household weight and the inverse of the probability of selection of the adult. The expression for the adult initial weight, $ADA0W_j$, is

$$ADA0W_j = ADCNT_i \cdot HHA7W_i,$$

where $ADCNT_i$ is the total number of adults in household *i*, and $HHA7W_i$ is the poststratified household weight.

4.2 Adult Nonresponse Adjustment

In some households the screener interview was completed but the sampled adult did not complete the extended adult interview. In addition, in few cases it was discovered that the sampled person was ineligible during the extended interview (i.e., the sampled person was not an adult or the sampled

[13] Completed adult extended interviews include partially completed adult interviews, provided the adult completed through Section I on health insurance.

[14] In CHIS 2001, in most households, the adult was sampled with equal probability from all enumerated adults except in those households that contained adults younger than 24 years old, adults 40 years or older, and no adults with unknown age. In these households, the probability of selecting adults 40 years old or older was two times the probability assigned to younger adults. This scheme reduced the chance of sampling adult children, thereby increasing the chance of including more children and adolescents in the survey since persons under 18 are linked to their parents in sampling.

adult from a race/ethnic sample was not a member of the target group) and should had not been sampled. 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. Before this adjustment, the extended interviews were classified into response groups as indicated in Table 4-1.

Table 4-1. Extended interview response status 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 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 computed as

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

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

The variables listed in Table 4-2 were considered in creating the nonresponse adjustment cells. A nonresponse analysis showed the response rates by sex and age groups were the most variable and best candidates for the cells; the nonresponse cells were created by classifying adults by sex and age groups within sampling strata. Cells with less than 30 respondents or with large adjustment factors were combined with adjacent cells. Sampling stratum was considered a hard boundary that was not crossed when the cells were combined. Appendix B, Table B-2 (rows 2.2a through 2.3c) shows the sum of weights before and after the nonresponse adjustment. Ineligible persons were dropped following this weighting step; further adjustments were performed only on records for eligible persons.

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

Variable	Levels
Sex of adult respondent	1. Male 2. Female
Presence of children and/or adolescents in the household at the screener level	1. Yes 2. No
Adult age group	1. 18-30 years old 2. 31-45 years old 3. 46-65 years old 4. 65 years or older
Household mail status	1. With a mailable address 2. Without a mailable address

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

4.3 Adult Trimming Factor

Before benchmarking the adult weights to known totals of adults in California in 2003, 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. Trimming reduces the weight and the influence of the observation on the estimates and their variances.

As in CHIS 2001, we computed statistics to identify influential weights that were candidates for trimming. These statistics and other variations were later studied in more 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., $z_i = w_{(i)},\ldots,w_{(i-1)}$). 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 in CHIS 2003 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

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

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

The three statistics for the largest 20 weights were examined separately for each sampling stratum. When these three statistics were each greater than 1 then the case was a candidate for trimming, but it was not necessarily trimmed. The final decision on trimming involved the inspection of the distribution of the weights within sampling stratum. Trimming occurred when the candidate weights appeared extreme within the distribution.

In CHIS 2003, 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* computed as

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

where $0 < t_i < 1$.

Forty-one adult records were trimmed. 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 examined again and new decisions were made about trimming. The raking adjustment might have changed the decision about which weights should be trimmed or the magnitude of the trimming factor. In such a case, the trimmed and raked weights were discarded and new trimming and weighting were undertaken. The number of trimmed weights reported here is at the completion of this process. The trimming factor, t_i, ranged from 0.67 to 0.8. Table B-2 in Appendix B (row 3.1 through 3.3) shows the strata with trimmed weights, and the sum of weights before and after trimming.

4.4 Adult Raking Adjustment

The final step in the adult weighting was raking the trimmed weights to population control totals to produce estimates consistent with the 2003 California Department of Finance (DOF) Population Projections. Included in the raking adjustment is a nontelephone adjustment discussed in section 4.5. The DOF provides population estimates at the county level by race, ethnicity, gender, and single age for each year from 2000 to 2050. The DOF files provide more up-to-date estimates than the Census 2000 Summary File 1 (SF1), and they can be used not only as the source of control totals for CHIS 2003 but also for future rounds of CHIS. The specific control totals and the method used to create them are described in Chapter 7.

Raking is a commonly used adjustment 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 of control totals (a dimension) then these adjusted weights are poststratified to another dimension (another set of control totals). 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

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

where $RAKEDF_{km}$ is the raking factor for dimension k, level l which adult i is in. For example, if 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 $RAKEDF_{4_1}$ if the adult i is male. The raking factors are derived so that 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. Table B-2 in Appendix B (rows 3.3 and 4.2) shows sum of weights before and after the raking adjustment.

4.5 Nontelephone Adjustment

Since CHIS 2003 was a traditional RDD sample, households without a landline telephone and households with only cellular telephones did not have a chance of being selected. The resulting undercoverage bias is related to the percentage of households without landline telephones and the differences in characteristics of the persons residing in households with a landline and households with a cellular telephone only or no telephone. In the 2000 Census approximately 1.5 percent of households in California did not have a telephone (U.S. Census Bureau, 2001). (It is not clear whether cell-only households are nontelephone or telephone households in this estimate.) Recent studies (Ford 1998; Anderson, Nelson, and Wilson 1998) show that the health characteristics of those with and without telephones are not as different as they had been in the past. Based on these factors, it is unlikely that most estimates from CHIS 2003 will have substantial bias as a result of the exclusion of households without a telephone. However, some estimates that are very directly correlated to income may be subject to greater biases due to this form of undercoverage. Although the percentage of households without a landline telephone was still very small in California in 2003, a special weighting adjustment was used to reduce this bias due to the excluding these households. Similar to its implementation in CHIS 2001, this adjustment was included as an additional raking dimension at the person level. The adjustment could not be done at the household level because households with no adult completed interview did not have the required data for the adjustment.

RDD surveys only sample from telephone exchanges devoted to landline service, and households with cellular telephones only are not covered. As the popularity of cellular telephones increases, and as the rate of cellular telephone-only households increases, the undercoverage associated with cellular telephones becomes more important because it also increases (see Tucker et al., 2004). In addition, initial studies of the characteristics of persons in cellular telephone only households indicate these persons differ from those households with landlines even with respect to health characteristics such as insurance (see Blumberg et al., 2004). In CHIS 2003 as well as in CHIS 2001, no weighting adjustment was developed specifically to account for households with only cellular telephones; however, this source of bias may need to be addressed in future rounds of CHIS.

The type of adjustment used in CHIS 2003 was based on the findings from CHIS 2001. In CHIS 2001, two adjustment methods were explored. The first method, called the Keeter adjustment (Keeter, 1995 and Brick, Waksberg and Keeter, 1996), adjusts the weights of persons in sampled telephone households who have had telephone service interruptions during a predetermined time period before the interview. Although the Keeter adjustment was considered for CHIS 2001, it was not used because very few adults lived in households with interruptions in telephone service in the sample. If this

type of adjustment had been implemented, the weights of the few records with interruptions would have increased so much that they would have had an undue influence on the final estimates and distribution of the weights.

The second method for adjusting for nontelephone households that was explored and eventually implemented in CHIS 2001 was a variation of the calibration method proposed by Ferraro and Brick (2001). In this method, calibration cells are created explicitly for adjusting weights for households without a telephone. Logistic regression or another similar procedure is used to compute the propensity of being a telephone household using an external file that includes all households. Cells that are homogeneous with respect to telephone status are formed by grouping cases with similar propensities. Using the same model, these calibration cells are recreated in the CHIS sample file. The external file is used to create the control totals since it includes records for all households. This approach has two limitations. First, the survey and the external control files must both contain all the variables that are used to compute the predicted propensities. The second limitation is related to the consistency of the estimates between the survey file and the control file. Since the cells are created using common variables, it is important to ensure that the variables are measured consistently. This same limitation applies to all poststratification or calibration weighting to external control totals.

The calibration method used in CHIS 2001 was also implemented in CHIS 2003. Because the telephone interruptions questions were not included in the CHIS 2003 questionnaire, we could not evaluate the feasibility of implementing a Keeter-type adjustment. For CHIS 2003, the 2002 American Community Survey Public Use Microdata Sample (ACS-PUMS) served as the external file to create control totals for the nontelephone adjustment. The 2002 ACS-PUMS included persons in telephone and nontelephone households and was large enough to produce reliable estimates for California. In contrast, in CHIS 2001 the March 2000 Current Population Survey (CPS) was used as the external file. When the 2002 ACS-PUMS data became available we decided to use that file as the external file for CHIS 2003. The 2002 ACS reports 1.4 percent of persons in households without telephone service in California[15]. This percentage is consistent with the Census 2000 and is considerably smaller than the percentage reported in the CPS. However, for the calibration method used in CHIS 2003, this estimate is not used when deriving the control totals.

Table 4-3 shows the set of variables that are common to both CHIS 2003 and the 2002 ACS-PUMS. We tabulated various estimates to verify that both surveys produce consistent

[15] The estimate of the percentage without a telephone service used the variable TEL from the ACS data files. The question for this variable is, "Is there telephone service available in this house, apartment, or mobile home from which you can both make and receive calls?" It is not clear whether cell-only households are nontelephone or telephone households in this estimate.

estimates. Only variables that produced similar estimates[16] were considered as predictors. The estimates for the CHIS sample were computed using the CHIS 2003 raked weights[17] including all dimensions except the nontelephone dimension.

Table 4-3. Common variables between CHIS 2003 and the 2002 ACS PUMS

Variable for consideration	Variable description	Variables used in cell creation
SEX	Sex	
RACE_ETH	Race / ethnicity	
EDUCATION	Education level of responding adult	✓
PUBLASSIST	Household receives governmental public assistance (SSI/AFDC…)	
MSTATUS	Marital status	
TENURE	Household rented or owned	✓
NADULT	Number of adults in the household	✓
NCHILD	Number of children in the household	
NTEEN	Number of teens in the household	

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

The race/ethnicity, educational attainment, and household tenure variables in the CHIS 2003 sample had missing values due to item nonresponse. These missing values were replaced using hot-deck imputation prior to comparing estimates. Hot-deck imputation is a technique where cases with missing values for specific variables are filled in with values from donor cases. See Chapter 8 for additional details of the hot-deck imputation and how these variables were imputed.

Using the variables in Table 4-3, we created the calibration cells for CHIS 2003. The goal was to create cells within which all households had a similar propensity of having a telephone. We used the categorical search algorithm CHAID (Kass, 1980) with the 2002 ACS-PUMS data instead of logistic regression used in Ferraro and Brick (2001), where the dependent variable was the telephone status (i.e., telephone household, nontelephone household). CHAID divides the data into groups in a stepwise fashion so that the propensities between the cells are as different as possible. Through a series of chi-square tests for equality of distributions, CHAID identifies the most important predictors and splits the data set into categories. Each of these categories is further segmented based on other predictors. The merging and splitting continues until no more statistically significant predictors are found or until a user-specified stopping rule is met. Using CHAID has two advantages over the logistic regression approach. First, the

[16] In some cases, differences in the estimates might be resolved by bringing in additional variables to make the estimates more consistent. However, for this purpose we wished to use only those variables that were most directly comparable.

[17] For CHIS 2001, a set of weights for the CHIS July 2000 RDD sample was created following the same adjustments as described as in this report except for the nontelephone adjustments. The July 2000 sample was about one-third of the full 2001 RDD sample and was used for preliminary analysis. No such preliminary set of weights was available for CHIS 2003.

interactions among the predictors are easily identified. Second, there is no need to group records with similar telephone propensities because the cells are created in the CHAID analysis. The final cells were created by collapsing the CHAID cells so there were 100 or more respondents in each cell.

Household tenure, race/ethnicity, educational attainment, and number of adults in the household were among the strongest predictors of telephone propensity. Race/ethnicity was not included in the model used to create the cells because other raking dimensions controlled for race/ethnicity. After the CHAID analysis on the ACS-PUMS, the same cells were created in the CHIS 2003 sample. The control totals were derived for the same cells using the 2003 California DOF Population Projections and the 2002 ACS-PUMS. Table 4-4 shows the definition of the cells used for the nontelephone adjustment in CHIS 2003.

Table 4-4. Nontelephone adjustment cell definition for CHIS 2003

Cell	Household tenure	Number of adults in the household	Educational attainment
1	Own		
2	Rent	1 adult	Less than 25 years old
3	Rent	1 adult	25 years or older, less than HS education
4	Rent	1 adult	25 years or older, High School grad or GED recipient
5	Rent	1 adult	25 years or older, At least some college
6	Rent	2 or more adults	

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

5. CHILD WEIGHTING

Final child weights were created for all completed child extended interviews[18] in CHIS 2003. For CHIS 2003 a child is defined as a person younger than 12 years old. The steps for the child weighting are similar to those described in the previous chapter for adults. One exception is that an additional weighting adjustment is needed to account for sampled adults who do not respond, because children are selected in the adult interview. A complete discussion of this adjustment is given in Section 5.1. The remainder of the chapter follows the same approach used for the adult weighting, with the creation of the child initial weights and the adjustments for nonresponse, trimming, and raking.

5.1 Household-Level Adjustment

The main difference between the child and adolescent weighting procedures and those of the adults is that the adults were sampled in the screener and persons under 18 years of age were sampled in the adult interview. Consequently, if an adult was sampled but not interviewed, the child and adolescent weights must be further adjusted to account for this level of nonresponse at the adult interview level.

Children and adolescents were sampled in Section G of the adult extended interview. 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 $HHA7W_i$ was adjusted. We refer to this adjusted weight as the Section G adjusted household weight, $HHA8W_i$, and it is

$$HHA8W_i = HHA8F_c * HHA7W_i ,$$

where,

$$HHA8F_c = \begin{cases} \dfrac{\sum_{i \in SG\_C, SG\_NC} HHA7W_i \cdot \delta_i(c)}{\sum_{i \in SG\_C} HHA7W_i \cdot \delta_i(c)} & i \in SG\_C \\ 0 & i \in SG\_NC \end{cases}$$

and SG_C is the set of adults who completed Section G, SG_NC is the set of adults who did not complete Section G, c denotes the Section G adjustment cell, and $\delta_i(c)$ is 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

[18] The adult who is most knowledgeable (MKA) about the child was interviewed to obtain the data.

who completed section G and were considered either completed or partially completed.[19] Note that this adjustment can be considered as a household adjustment on top of the poststratified household weight.

The Section G nonresponse adjustment cells were created within sampling stratum using a combination of the telephone mail 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. Table B-1 in Appendix B (rows 9.1 through 9.2) shows the sum of weights before and after the section G nonresponse adjustment.

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. In CHIS 2003, 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 adult (i.e., the sampled adult is the parent or legal guardian of the child). If the sampled adult did not have a child associated with him or her, then no child was sampled even if there were children present in the household.

Since the child sampling depends on the associations of children and adults within the household, these associations were defined before sampling children. In CHIS 2001, children who were not associated with any adult in the household were randomly linked to an adult. In CHIS 2003 the procedure was revised and only children who had a parent or legal guardian in the household were eligible to be sampled. (See Report 1: Sample Design for information on the within-household person selection process.)

The probability of selection should reflect 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[20] of the sampled child. Accordingly, the initial child weight, $CHA0W_j$, is

$$CHA0W_j = \frac{1}{CHPROB_j} \cdot HHA8W_i$$

[19] If the adult interview was not completed no attempt was made to complete a child/teen interview.

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

where $CHPROB_j$ is the probability of selecting the *j*-th child associated to the *i*-th sampled adult computed as

$$CHPROB_j = \begin{cases} \dfrac{1}{ADLTCNT_i} \cdot \dfrac{1}{SACHCNT_i} & \text{If the sample adult does not have a spouse/partner living in the household or if the spouse/partner of the sample adult is not the parent or legal guardian of the sample child} \\ \dfrac{1}{ADLTCNT_i}\left(\dfrac{1}{SACHCNT_i} + \dfrac{1}{SPCHCNT_i}\right) & \text{If the sample adult has a spouse/partner living in the household and the spouse/partner of the sample adult is the parent or legal guardian of the sample child} \end{cases}$$

where $ADLTCNT_i$ is the number of adults in the household, $SACHCNT_i$ is the number of children for whom the sampled adult is the parent or legal guardian, $SPCHCNT_i$ is the number of children for whom the spouse/partner is the parent or legal guardian. The number of sampled children and sum of the initial weights are in Table B-3 in Appendix B (rows 1.1 and 1.2).

5.3 Other Child Weighting Adjustment

As mentioned before, the adjustments made to the child weights are the same as the adjustments to the adult weights. These adjustments are for extended interview nonresponse, trimming influential weights, and raking to control totals. The raking adjustment includes a dimension to account for children living in nontelephone households.

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 mail status, sex, and child 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 mail status, sex, and age group at the state level to determine the variables with the most variable response rates. Using these results, the cells are defined by sampling strata, sex, and age group. Any cells still containing fewer than 30 respondents were collapsed over age group. The smallest

two strata were collapsed across both sex and age group. Table B-3 in Appendix B (rows 2.1 to 2.3c) shows the number of sample records and sum of weights before and after the nonresponse adjustments.

The next step was to identify and trim influential 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 53 weights in the child component of CHIS 2003. The trimming factors range from 0.40 to 0.80. Table B-3 in Appendix B (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 2003 California Department of Finance Population Projection results. See Chapter 7 for the specific controls used. The expression for the raking adjustment is the same as the one for adult weights described in Section 4.4. Table B-3 in Appendix B (rows 3.3 and 4.2) shows sum of weights before and after the raking adjustments.

6. ADOLESCENT WEIGHTING

In CHIS 2003 adolescents were sampled and responded to the interview for themselves after a parent or guardian gave the interviewer permission to conduct the interview. In CHIS 2003, adolescents are defined as persons between the ages of 12 and 17, inclusive. In this section we describe the creation of analytic weights for the adolescent interview.

6.1 Adolescent Initial Weight

The procedures for creating the adolescent weights are the same as the procedures for creating the child weights described in Chapter 5. As with the child weighting, the initial weights for the adolescent incorporates the probability of sampling the adult and the probability of sampling an adolescent among the adolescents associated with the sampled adult. The initial weight, $TNA0W_j$, is

$$TNA0W_j = \frac{1}{TNPROB_j} \cdot HHA8W_i$$

where $HHA8W_i$ is defined in chapter 5 as the nonresponse adjusted adult weight and $TNPROB_j$ is computed the same as $CHPROB_j$ in chapter 5. The number of adolescents sampled and initial adolescent weights are in Table B-4 in Appendix B (rows 1.1 and 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. Appendix B, Table B-4 (row 2.3) shows the nonresponse adjusted adolescent weight. Initially the adolescent nonresponse adjustment cells were created using household mail status, sex, and adolescent 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 the cells using sampling strata, sex, and age group. Cells still containing fewer than 30 respondents were collapsed across sex first and then age group.

After the nonresponse adjustments, 63 influential weights were identified and trimmed. Table B-4 in Appendix B (rows 3.1 through 3.3) gives the trimmed weights by self-reported strata and the sum of the weights before and after applying the trimming factors to the adolescent weights.

The last step of weighting was to rake the weights to the control totals. The adolescent weights were raked to population control totals to produce estimates consistent with the 2003 California DOF Population Projection results. See Chapter 7 for details on the control totals. The expression for the raking adjustment is exactly the same as used in the raking of the adult weights and the child weights. Table B-4 in Appendix B (rows 3.3 and 4.2) shows the sum of weights before and after the raking adjustments.

7. RAKING AND CONTROL TOTALS

This chapter describes the raking procedure in the weighting and the development of control totals for the CHIS 2003 sample. The first section gives a general overview of raking and why this procedure was used in CHIS 2003. The second section describes the 11 dimensions used to rake the 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 for the dimensions were derived from the 2003 California Department of Finance Population Projection 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 for more auxiliary information. The limitation in poststratification arises because each unit falls into only one cell and the number of respondents in a cell should not be small. With raking, the cell size is based on the distribution of each raking dimension. For example, if poststratification were used, only some cross-classified age/race/sex categories could be used in the adjustments, whereas, with raking, more levels of these variables without the full cross-classification and important geographic level data such as county can also be included as dimensions. As we mentioned in Section 4.4, raking can be thought of as a multidimensional poststratification procedure, because the weights are basically poststratified to one set of control totals (a dimension), 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 (at least within a specified tolerance). Raking was also used in CHIS 2001. Below, we describe the procedure in more detail. Brackstone and Rao (1979); Deville and Särndal (1992); and Kalton and Flores Cervantes (2003) also describe aspects of 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 becomes highly likely that poststratification produces unstable estimates unless the cells in the cross-tabulation are collapsed. With the 11 dimensions used in CHIS 2003, the potential for collapsing would be very extensive. An alternative estimation approach is to rake the weights to the marginal totals of the counts. 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.

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 only 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 crossing of the variables. This means that deficient cells (cells with small sample sizes) are defined by looking at the sample sizes of the marginals. Furthermore, this permits the use of more variables or control totals than is possible with poststratification.

7.2 Raking Dimensions

The 11 dimensions used in CHIS 2003 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. Section 4.5 has more details on the nontelephone adjustment and the variables used to create the levels for this dimension. For reference, there were also 11 dimensions in CHIS 2001. Although many of the CHIS 2001 dimensions are similar to the dimensions used in CHIS 2003, there are significant differences, especially in the definitions of the cells that use race.

In CHIS 2003, although the adult, adolescent, and child weights were adjusted separately for nonresponse, the weights were not raked separately, but rather in a single file. The change was necessary because there were no available separate control totals for adults as in CHIS 2001. A different file was used as a source of the control totals in CHIS 2003. One consequence of this change was that the number of iterations required for convergence increased, mainly due to the increased complexity and size of the combined file.

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
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

See note at end of table.

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

Dimension	Level	Description	Categories	
4	SPAs in Los Angeles Co., Alameda County, Remainder of CA	SPAs (8), Alameda Co. (3), Remainder of CA (1)	1	SPA 1
			2	SPA 2
			…	…
			7	SPA 7
			8	SPA 8
			9	Hayward Census Place
			10	Oakland Census Place
			11	Remainder of Alameda County
			12	Remainder of CA
5	Region (collapsed where necessary)	Race/ethnicity (7)	1	Hispanic
			2	White not Hispanic
			3	African American not Hispanic
			4	American Indian not Hispanic
			5	Asian not Hispanic
			6	Native Hawaiian not Hispanic
			7	Two or more races not Hispanic
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, White non-Latino, under 12 years
			122	Male, White non-Latino, 12 to 17 years
			123	Male, White non-Latino, 18 to 64 years
			124	Male, White non-Latino, 65 years or older
			131	Male, African American non-Latino, under 12 years
			132	Male, African American non-Latino, 12 to 17 years
			133	Male, African American non-Latino, 18 to 64 years
			134	Male, African American non-Latino, 65 years or older
			141	Male, American Indian non-Latino, under 12 years
			142	Male, American Indian non-Latino, 12 to 17 years
			143	Male, American Indian non-Latino, 18 to 64 years
			144	Male, American Indian non-Latino, 65 years or older
			151	Male, Asian non-Latino, under 12 years
			152	Male, Asian non-Latino, 12 to 17 years
			153	Male, Asian non-Latino, 18 to 64 years
			154	Male, Asian non-Latino, 65 years or older
			161	Male, Native Hawaiian non-Latino, under 12 years
			162	Male, Native Hawaiian non-Latino, 12 to 17 years
			163	Male, Native Hawaiian non-Latino, 18 to 64 years
			164	Male, Native Hawaiian non-Latino, 65 years or older
			171	Male, Two or more races non-Latino, under 12 years
			172	Male, Two or more races non-Latino, 12 to 17 years
			173	Male, Two or more races non-Latino, 18 to 64 years
			174	Male, Two or more races non-Latino, 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	
			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, White non-Latino, under 12 years
			222	Female, White non-Latino, 12 to 17 years
			223	Female, White non-Latino, 18 to 64 years
			224	Female, White non-Latino, 65 years or older
			231	Female, African American non-Latino, under 12 years
			232	Female, African American non-Latino, 12 to 17 years
			233	Female, African American non-Latino, 18 to 64 years
			234	Female, African American non-Latino, 65 years or older
			241	Female, American Indian non-Latino, under 12 years
			242	Female, American Indian non-Latino, 12 to 17 years
			243	Female, American Indian non-Latino, 18 to 64 years
			244	Female, American Indian non-Latino, 65 years or older
			251	Female, Asian non-Latino, under 12 years
			252	Female, Asian non-Latino, 12 to 17 years
			253	Female, Asian non-Latino, 18 to 64 years
			254	Female, Asian non-Latino, 65 years or older
			261	Female, Native Hawaiian non-Latino, under 12 years
			262	Female, Native Hawaiian non-Latino, 12 to 17 years
			263	Female, Native Hawaiian non-Latino, 18 to 64 years
			264	Female, Native Hawaiian non-Latino, 65 years or older
			271	Female, Two or more races non-Latino, under 12 years
			272	Female, Two or more races non-Latino, 12 to 17 years
			273	Female, Two or more races non-Latino, 18 to 64 years
			274	Female, Two or more races non-Latino, 65 years or older
7	State	Asian groups (5) x Age groups (4) (collapsed where necessary)	11	Non-Hispanic Chinese only, under 12 years
			12	Non-Hispanic Chinese only, 12 to 17 years
			13	Non-Hispanic Chinese only, 18 to 64 years
			14	Non-Hispanic Chinese only, 65 years or older
			21	Non-Hispanic Korean only, under 12 years
			22	Non-Hispanic Korean only, 12 to 17 years
			23	Non-Hispanic Korean only, 18 to 64 years
			24	Non-Hispanic Korean only, 65 years or older
			31	Non-Hispanic Filipino only, under 12 years
			32	Non-Hispanic Filipino only, 12 to 17 years
			33	Non-Hispanic Filipino only, 18 to 64 years
			34	Non-Hispanic Filipino only, 65 years or older
			41	Non-Hispanic Vietnamese only, under 12 years
			42	Non-Hispanic Vietnamese only, 12 to 17 years
			43	Non-Hispanic Vietnamese only, 18 to 64 years
			44	Non-Hispanic 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	Hispanic, under 12 years
			12	Hispanic, 12 to 17 years
			13	Hispanic 18 to 64 years
			14	Hispanic 65 years or older
			21	White not Hispanic, under 12 years
			22	White not Hispanic, 12 to 17 years
			23	White not Hispanic, 18 to 64 years
			24	White not Hispanic, 65 years or older
			31	Non-White not Hispanic, under 12 years
			32	Non-White not Hispanic, 12 to 17 years
			33	Non-White not Hispanic, 18 to 64 years
			34	Non-White not Hispanic, 65 years or older
9	State	Education (4)	1	Not applicable (age < 25 years)
			2	Less than High School
			3	High School grad or GED recipient
			4	At least some college
10	State	# Adults in HH (3)	1	0 or 1 adult,
			2	2 adults,
			3	3 or more adults
11	State	Non-telephone	1	Homeowner
			2	Renter, 1 adult in household, less than 25 years old
			3	Renter, 1 adult in household, 25 years old or older, less than HS
			4	Renter, 1 adult in household, 25 years old or older, High School grad or GED recipient
			5	Renter, 1 adult in household, 25 years old or older, At least some college
			6	Renter, 2 or more adults in household

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

Before raking, dimensions with levels or cells with fewer than 50 respondents were collapsed with "adjacent" cells. Since raking was used, a larger number of respondents was needed than what is typically required for poststratification. In dimensions 1, 2, 5, and 8 the collapsed cells were created by combining counties within the geographic regions shown in Table 7-2. Dimension 5 was initially defined at the region level, and some cells were collapsed across regions because some 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 and 10 were defined at the state level because the information needed to create these cells (education and type of household defined by number of adults in the household) was not available at the county level. When collapsing the cells, we ensured that there was at least one marginal total that preserved the number of persons in the county/self-reported stratum. In this way, the raked weights summed to the control total for each stratum.

Table 7-2. Regions in California

Region	Counties
Northern & Sierra Counties	Butte, Shasta, Del Norte, Lassen, Humboldt, Modoc, Siskiyou, Trinity, Lake, Mendocino, Colusa, Glenn, Tehama, Sutter, Yuba, Nevada, Plumas, Sierra, 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, 2003 California Health Interview Survey.

Table 7-3 shows the overall adjustment factors for the adult, child, and adolescent weights. The overall adjustment factors were computed as the ratio of the control total to the sum of weights before raking. These factors are, in some sense, a measure of the magnitude of the bias correction for estimates of totals. Since the weights were already adjusted for nonresponse, the raking factor could be used as an indirect measure of undercoverage. This interpretation was not straightforward in CHIS 2003 because the weights were adjusted at the household level for the 2001 total number of households with and without individuals under 18 years old (see Section 3.8). The adjustment factors also confound several factors such as reporting error and residual nonresponse error. Nevertheless, they may be used as a rough indicator of relative within-household coverage error. A factor greater than unity suggests undercoverage, and a factor less than unity suggests overcoverage (these are all relative measures).

Table 7-3. Overall adjustment raking factors for adult, child, and adolescent interviews by sample characteristics

Characteristic	Adult	Child	Adolescent
Total	1.118	1.074	1.068
Sex			
Male	1.151	1.084	1.080
Female	1.088	1.064	1.055
Age group			
Under 5 years		1.080	
6 – 11 years		1.069	
12 – 17 years			1.068
18-24 years	1.241		
25-29 years	1.285		
30-39 years	1.242		
40-49 years	1.098		
50-64 years	0.980		
65 years and over	1.021		
Race/Ethnicity[a]			
Latino	1.316	1.108	1.104
Non-Latino			
White alone, non-Latino	0.990	0.929	0.939
African American alone, non-Latino	1.388	1.610	1.988
American Indian/Alaska Native alone, non-Latino	0.968	1.525	0.854
Asian alone, non-Latino	1.314	1.358	1.525
Native Hawaiian, Pacific Islander alone, non-Latino	1.102	1.542	0.762
Two or more races, non-Latino	0.650	0.845	0.538
Asian ethnic groups (non-Latino)[a]			
Chinese only	1.316	1.108	1.104
Korean only	1.183	1.431	1.870
Filipino only	1.254	1.244	1.508
Vietnamese only	1.670	1.438	1.252
Educational Attainment			
Not applicable (age < 25 years)	1.241	1.074	1.068
Less than High School,	1.677		
High School grad or GED recipient,	1.026		
At least some college	1.003		
Household Tenure[b]			
Owner	1.003	0.986	0.971
Renter	1.348	1.197	1.262

See notes at end of table.

Table 7-3. Overall adjustment raking factors for adult, child, and adolescent interviews by sample characteristics (Continued)

Characteristic	Adult	Child	Adolescent
Number of adults in the household[b]			
One	1.101	1.241	1.250
Two	1.117	1.071	1.058
Three or more	1.127	1.017	0.993
Number of children in the household[b]			
None	1.093		1.055
One	1.172	1.096	1.085
Two or more	1.240	1.065	1.084
Number of adolescents in the household[b]			
None	1.115	1.076	
One	1.144	1.060	1.082
Two or more	1.149	1.073	1.053

[a] OMB definition
[b] Person level estimate by type of household.
Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

Table 7-3 shows that the adult, child, and adolescent undercoverage is minimal. For adults, the adjustment factor for males is larger than for females, which is common in household surveys and was the case in CHIS 2001. The factors also indicate a potential for undercoverage for younger adults.

One large overall adjustment factor is for persons who self-reported as having less than a high school education (1.677). The factors for the Asian and Non-Latino African American ethnic groups are also all larger than unity suggesting potential undercoverage. Other factors worth noting include those for persons who self reported as being two or more races. The factors for these race groups are much less than unity, suggesting the CHIS 2003 estimates of persons of two or more races before raking are much higher than the corresponding DOF 2003 totals. Although the reason has not been determined, it seems likely that the way the race question was asked in CHIS (prompting for "any other race?") encouraged the reporting of multiple races as compared to the Census form. Note that the race-ethnicity overall adjustment factors were computed using the OMB definition (see Chapter 8) where the number of persons who reported two races or more is smaller than if the Census definition were used.

7.3 Sources Used to Produce the Control Totals for CHIS 2003

Several sources were considered for control totals for CHIS 2003. The sources include the Census 2000 files (U.S. Census Bureau, 2001), the 2002 American Community Survey (ACS) (U.S. Census Bureau, American Community Survey, 2002), and the 2003 California DOF Population Projection (State of California, Department of Finance, 2004).

Based on recommendations from UCLA, the 2003 California DOF Population Projections were used as the main source for the demographic control totals, (i.e., raking dimensions defined by gender, race, ethnicity, age, and stratum). These population totals had to be adjusted to remove the population living in group quarters who are not included in CHIS. The 2000 Census files were used to compute the proportion of persons living in group quarters. The 2002 ACS files were used as a source for educational attainment, household tenure, and household composition. The following paragraphs describe each source and how they were used to create raking dimension control totals for CHIS 2003.

7.3.1 California Department of Finance Population Projection

The main source used for creating raking dimensions is the race-ethnicity population projection published by the California DOF. The DOF provides population estimates at the county level by race, ethnicity, gender and single age for each year from 2000 to 2050. The DOF population projections use the 2000 Census counts not adjusted for the Census 2000 undercount as the baseline. The DOF used a baseline cohort-component method 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 characteristics and behavior 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 2003. A significant advantage of using the DOF files is that they provide the most up-to-date estimates and they can be used not only as the source of control totals for CHIS 2003 but also for future rounds of CHIS.

The DOF population projection files[21] contain counts for each county in California by year of age (101)× sex (2)× race (6)× ethnicity (2) as defined in Table 7-4. The DOF has only one multi-race category that includes persons with two or more races.

Table 7-4. Definition of counts available in the 2000 and 2003 California DOF population projection files*

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

* Available at the county level
Source: State of California, Department of Finance.

The main disadvantage of the DOF projections is the definition of race on the file. The DOF population projections use the U.S. Office of Management and Budget (OMB) race definition, also known as "modified" race (there are no separate population counts for "other" race). The DOF projections 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 (both in 2001 and 2003), respondents who could not identify themselves as any of the five OMB race categories could answer with a sixth category, "some other race." This is consistent with the 2000 Census method. Recoding of "other race" for CHIS 2003 largely followed Census procedures as well (see Report 3: Data Preparation).

[21]The file that included race group breakdowns for those of Hispanic ethnicity was provided upon request by the California DOF. The public use file only contains race group breakdowns for non-Latinos.

In order to use the DOF projections, any sampled person who reported other race (alone or in combination with another race) had to be recoded into one or more of the OMB categories. This would have required the imputation of an OMB race category for 6,642 persons (12.2 percent) who self-reported "other race" only in the CHIS 2003 sample. As an alternative, a variable that combined ethnicity with the OMB race that reduced the number of imputations was proposed and approved by UCLA for CHIS 2003. The recoding includes an additional level that groups Latinos of any race as shown in Table 7-5. Because most respondents who self-reported other race only were Latinos, the number of imputed records was reduced significantly to 128 persons (0.2 percent) who self-reported 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 7-5. 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

Another difficulty associated with the use of DOF population estimates is that they include the population living in group quarters. Since the eligible population in CHIS 2003 excludes persons in group quarters, these persons were removed from the DOF population estimates. The Census 2000 files were used to estimate the proportion of persons in group quarters and these proportions were applied to the DOF projections. At the time control totals were being developed, the DOF did not have separate projections for the population living in group quarters. For more information about how the group quarters populations were removed from the DOF projections, see Section 7.4.

7.3.2 Census 2000 Files

The Census 2000 Summary Files 1 and 2 (SF1and SF2) were originally used to derive household and person control totals in CHIS 2001. Additional Census files with information from the long form not available during CHIS 2001 were available for use as control totals for CHIS 2003. These

files have socio-economic characteristics not generally found in other sources that are restricted mainly to geographic and demographic information. The Census 2000 files can also be used to produce population counts by geographic areas other than counties (i.e., Census Tracts and Census Places).

When considering the use of use the SF1 and SF2 files as the main source for control totals, a concern was that the counts would have to be inflated to the 2003 population level. The inflated totals would not reflect any differential growth at the stratum level between 2000 and 2003, nor would they account for differential growth by racial and ethnic groups. Due to this problem, and subsequent surveys likely having additional differential growth not accounted for, it was decided that the SF1 and SF2 files would not be used as the main source for control totals.

The Census files were used to derive the control totals for the dimension defined by SPAs in Los Angeles and cities in Alameda 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 2003[22]. Los Angeles SPAs were defined in terms of Census Tracts; Hayward and Oakland were defined in terms of Census Places.

7.3.3 American Community Survey for California

The American Community Survey (ACS) is a new nationwide survey providing current and detailed information for demographic, social, economic, and housing data. It will replace the decennial census long form in future censuses and is a critical element in the Census Bureau's reengineered 2010 Census plan. The 2002 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 2002 ACS public use micro level file (ACS-PUMS) for California that provides household and population estimates at the state level. For selected counties and large communities, previously generated tables for a subset of estimates can be downloaded from the Census website[23]. The 2002 ACS includes population estimates for 24 strata, plus Oakland City, but not for the SPAs in Los Angeles County, Hayward in Alameda County, and some small counties.

When sources for control totals were being evaluated for CHIS 2003, we analyzed the possibility of using the 2002 ACS as the source for household control totals in the household

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

[23] http://www.census.gov/acs/www/

poststratification adjustment (see Section 3.8). The 2002 ACS files provided household counts by the presence of a person under 18 years old for selected counties and for the state for 2002. In order to estimate the household growth in the missing counties, we modeled the household growth in these counties using overall population growth from 2000 to 2002 measured by the difference between the Census 2000 county population counts and 2002 ACS county population estimates. However, the variability in the predicted 2002 household counts was too large to make the predictions useful. We expected to have similar problems if the same method were used to produce population estimates for the missing counties. As a result, the 2002 ACS was rejected in favor of the 2003 DOF population projections for control totals in CHIS 2003.

Although the 2002 ACS was not used as the primary source of population control totals in CHIS 2003, it was used to compute proportions by educational attainment and type of household (tenure and number of adults in the household) that were later applied to the to 2003 DOF totals to derive the controls for the raking dimensions defined by these characteristics (dimensions 9, 10, and 11 in Table 7-1). These variables were not available in the DOF files. Applying the 2002 factors assumed that there were no changes in the proportions of the population between 2002 and 2003[24] for these variables.

7.4 Producing the Control Totals for CHIS 2003

The derivation of the control totals was a challenging task in CHIS 2003. 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 2003, there were 11 raking dimensions. Deriving the control totals independently could lead to inconsistencies between totals across the dimensions (or combination of levels of the dimensions). To overcome these difficulties, we developed a procedure in which most of the dimensions could be 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 and levels available from the sources. 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 among the dimensions. The details of this procedure are described in the following sections.

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

In CHIS 2001 and CHIS 2003, the population living in group quarters was not eligible for the survey. Therefore, 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 if counts of persons in group quarters for all variables and geographic levels are available. However, this information was not available in the DOF files. To compute the number, it was assumed that the proportion of the population in group quarters did not change between 2000 and 2003. As a result, the Census 2000 SF1 file could be used to compute these proportions. Note that this assumption is consistent with the assumption made by the California DOF for its population projections.

There were two problems when computing the percentage of the population living in group quarters using the Census SF1 file. The first problem was the limited number of group quarter counts that can be produced from the SF1 file. The only counts available are by stratum (41) ×age group (3) ×sex (2) × race (7) and by stratum (41) × age group (3) × sex (2) × ethnicity (3) as defined in Table 7-6. The file could not be used to produce population counts by single age or by the cross-tabulation of race and ethnicity.

The second problem was that the group quarter population counts from the SF1 file are defined for the seven race categories shown in Table 7-6 and not the five OMB race groups used in the DOF file (see Table 7-3). To address these problems, two assumptions were made. The first assumption was that the distribution of the population in group quarters was uniform among the age groups. For example, if the percentage of persons 18 to 64 years old in group quarters is 1.78 percent, then 1.78 percent of the persons 24 years old are assumed to be in group quarters. The second assumption was that the distribution of persons in group quarters by ethnicity (Latino or non-Latino) was also the same within race and age groups. For example, if 1.3 percent of the African American population is in group quarters, then 1.3 percent of the Latino African American is in group quarters. Using these assumptions, 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 20,664 cells denoted *rc*, where *r* denotes race and $c = \text{strata}(41) \times \text{ethnicity}(2) \times \text{age}(18) \times \text{gender}(2)$. 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 7-6.

Table 7-6. Definition of levels of variables for group quarters populations in the Census 2000 SF1 file

Characteristics	Available counts
Strata (41)	Counties or combinations of multiple counties defined in CHIS 2003
Age group (3)	Less than 18 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)	Hispanic or Latino
	White alone not Hispanic or Latino
	Other

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

The 18 levels of age (see Table 7-7) 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.

Table 7-7. 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

7.4.1 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 two 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 $m = \text{strata}(41) \times \text{race}_{\overline{OMB}}(7) \times \text{age group}(3) \times \text{sex}(2)$.

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 $n = \text{strata}(41) \times \text{ethnicity}(3) \times \text{age group}(3) \times \text{sex}(2)$ as in the SF1. Note that $D1_{m}^{2000\,GQ}$ and $D2_{n}^{2000\,GQ}$ 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 used in raking.

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 2003 within cell *rc*, the proportion $p_{rc}^{2000\,\overline{GQ}}$ could be used to compute $T_{rc}^{2003\,\overline{GQ}}$ defined as the 2003 total population not living in group in cell *rc*, as

$$T_{rc}^{2003\,\overline{GQ}} = p_{rc}^{2000\,\overline{GQ}} * T_{rc}^{2003},$$

where T_{rc}^{2003} is the 2003 total population from the 2003 California DOF file in cell *rc*. However, the total T_{rc}^{2003} was not available in the DOF file due to inconsistencies in racial categorization between the SF1 and the DOF projection. The 2003 population totals, $T_{sc}^{2003\,OMB}$, were only available in the DOF file for 17,712 cells (labeled *sc*) defined using the OMB racial categories. The cells *sc* were defined by the cross-tabulation $sc = \text{race}_{OMB}(6) \times c$, where the subscript *OMB* refers to the OMB race groups that exclude the category for "some other race" as shown in Table 7-8, and *c* is defined as before.

Table 7-8. 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}, expand by the individual population for the race groups as follows:

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

In the same way, the total population in a cell *c* can be separated by population totals 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 the OMB 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). The reallocation of the population between the cells *r* and *s* can be expressed mathematically.

In order to illustrate the reallocation, consider the Asian group (ignoring the strata, 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

$p_{AS\_TM}^{OMB}$ 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_{ASc}^{OMB}) is composed of the original non-OMB Asian alone total (T_{ASc}), 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 (omitting the subscript *c* for convenience) and T_O is defined as the group who reported "other race" only.

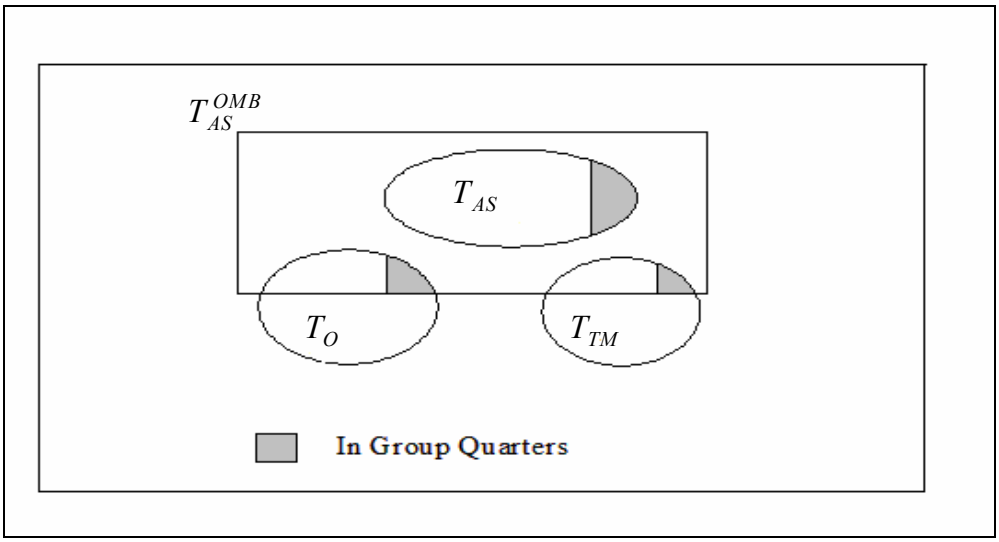


Figure 7-1. Relationship between OMB Asian alone and non-OMB groups

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 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 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 2003 the proportion was computed as

$$p_{sc}^{2003\,\overline{GQ}} = p_{sc}^{2000\,\overline{GQ}} = \frac{T_{rc}^{2000\,\overline{GQ}}}{T_{rc}^{2000}}.$$

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

$$T_{sc}^{2003\ OMB\ \overline{GQ}} = p_{sc}^{2000\ \overline{GQ}} * T_{sc}^{2003\ OMB} = \frac{T_{rc}^{2000\ \overline{GQ}} * T_{sc}^{2003\ OMB}}{T_{rc}^{2000}}.$$

where T_{rc}^{2000} is computed using the SF1 file, $T_{sc}^{2003\ OMB}$ using the 2003 DOF file and $T_{rc}^{2000\ \overline{GQ}}$ is the 2000 population in cell *rc* not in group quarters, as defined earlier. The 2003 total population not living in group quarters in California is computed as

$$T^{2003\ \overline{GQ}} = \sum_s \sum_c T_{sc}^{2003\ OMB\ \overline{GQ}}.$$

Table 7-9 shows the total population in the 2003 DOF file and the estimated total (and percentage) of the population living in group quarters.

Table 7-9. Population in California

Type	Population	%
In group quarters	849,423	2.4
Not in group quarters	35,085,544	97.6
Total	35,934,967	100.0

7.4.2 Computing the Control Totals

The totals $T_{sc}^{2003\ OMB\ \overline{GQ}}$ 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 totals $T_{sc}^{2003\ OMB\ \overline{GQ}}$ for Asian only. Using the Census 2000 SF1 files, we computed the percentages for the Asian groups in Table 7-10. The percentages of the Asian groups by ethnicity (Latino, non-Latino) were computed using the 2002 ACS-PUMS file because these counts were not available in the SF1 file. In this approach, it was assumed that there were no changes in the distribution of the Asian groups between 2000 and 2003.

Table 7-10. Census 2001 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 SPA in Los Angeles County and cities in Alameda County, used information from the 2000 SF1. The Los Angeles County Department of Health (LACDH) produced a listing of Census tracts by SPAs. 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 2003 DOF population total (excluding group quarters) to produce the controls for dimension 4. For cities in Alameda County (Hayward, and Oakland), the same SF1 file was used to compute the percentages of 2000 population of Alameda County in Hayward and Oakland. The percentages were applied to the 2003 DOF Alameda County population total (excluding group quarters). As in the previous dimension, it was assumed that there were no changes in the percentage of the population living in the SPAs and cities in Alameda County between 2000 and 2003.

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 2002 ACS-PUMS and then applied to the 2003 DOF population total (excluding group quarters). The underlying assumption was that there were no changes in the distribution of the population between 2002 and 2003.

8. IMPUTATION PROCEDURES

As is the case with any household survey, both unit and item nonresponse are virtually unavoidable. In this report we have described how weighting adjustments have been used to compensate for unit nonresponse in CHIS 2003. Report 4 discusses unit nonresponse in more detail. This chapter focuses on item nonresponse and the imputation for missing responses for certain variables. Only the imputation of variables used in weighting is covered here.[25] The imputed values were needed in the last stages of the weighting process and only interviews that were considered completed units had values imputed. The percentage of missing data and consequent imputation for virtually all of these items is small.

Section 8.1 describes the imputed variables (self-reported sex, age, household tenure, educational attainment, and race-ethnicity) and reviews the different types of imputation techniques used to fill in the missing data. The two imputation techniques employed in CHIS 2003 are random allocation, and hot-deck imputation. Section 8.2 reviews the imputation process for self-reported sex and age. A random allocation imputation process was used for both variables. Section 8.3 discusses the imputation of self-reported household tenure and educational attainment of the interviewed adult. Hot-deck imputation was used to impute these two variables. Section 8.4 covers imputation of race and ethnicity. The last section of this chapter, 8.5, lists the geographic location variables for CHIS 2003. 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 in CHIS 2003. Most of these variables were also imputed in CHIS 2001 except for household ownership status (or tenure), educational attainment, OMB self-reported race/ethnicity, and OMB self-reported non-Hispanic Asian group. Household tenure and educational attainment were included as components of additional raking dimensions not used in CHIS 2001. Race was a component of some raking dimensions in CHIS 2001, but was not defined in the same way as in CHIS 2003. 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

[25] Westat was responsible for imputing all the variables necessary for the weighting process except for the respondent's county of residence.

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	Extended interview items	Variable type
SRAGE	Self-reported age	AA2, CA3, TA2	Demographic
SRSEX	Self-reported sex	AA3, CA1, TA3	Demographic
SRTENR	Self-reported household tenure	AK25	Socio-economic
SREDUC	Self-reported educational attainment	AH47	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 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-Hispanic Asian group	AA5E_1- AA5E_18, TI7_1- TI 7_18, CH7_1-CH7_18	Race/ Ethnicity

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

In CHIS 2003 and CHIS 2001, random allocation and hot-deck imputation were used to fill in the missing responses. The first imputation technique is a completely random selection from the observed distribution. This method is used only when a very small percentage of the items are missing. For example, when imputing the missing values for self-reported age, the distributions of the responses for age by type of interview (adult, child, or adolescent) were used to randomly assign an age using probabilities associated with these distributions. More detail about this imputation process is given in Section 8.2.

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 CHIS 2003. The hot-deck approach is probably the most commonly used method for assigning values for missing responses in large-scale household surveys (Sande, 1983 and 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 2003, 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 randomly imputed from one of the donors in the pool. Once a donor is used, it is removed from the pool of donors.

8.2 Self-Reported Sex and Age

The percentages of cases where either sex or age was missing in CHIS 2003 are very small across all types of extended interviews (adult, child, and adolescent). The sex of only three children was imputed in CHIS 2003; no adults or adolescents were missing self-reported sex. The missing data for self-reported sex was imputed randomly. A (uniformly distributed) random number was generated for the three missing values. The value of the respective random numbers was each less than 0.50, therefore, the sex was assigned as female.

Age was imputed in 95 cases in CHIS 2003, whereas only 34 cases were imputed in CHIS 2001. Table 8-2 summarizes the number of cases that were imputed for age and gender in CHIS 2003. A similar procedure was used to impute the missing self-reported age values. If self-reported age was missing, it was determined if the respondent reported an age range (question AA2A in the adult interview). If no age range was reported then a random age value was drawn from the observed distribution of self-reported age, and the drawn age was used to replace the missing age. For example, assume a random number was assigned to a child and it had a value of 0.21. If the distribution of the ages of the children was such that 17 percent were age 4 or less and 23 percent were age 5 or less then the child's imputed age would be 5 years old. If the respondent did not report a specific age but did report an age range, then random draws were made from the observed distribution of self-reported age within the given range.

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

Person type	Number completed	Number missing sex	% missing sex	Number missing age	% missing age
Adult	42,044	0	0.00	84	0.20
Child	8,526	3	0.04	11	0.13
Adolescent	4,010	0	0.00	0	0.00
Total	54,580	3	0.01	95	0.17

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

8.3 Household Tenure and Educational Attainment

As described in Chapter 7, household tenure and the adult's educational attainment were used to create raking dimensions 9 and 11. Household tenure is a household-level variable collected in the adult interview with 434 missing tenure (1.04 percent). Education is a person-level variable with 182 (0.47 percent) with missing adult education attainment. Because control totals were available only for adults 25 years old or older, education was only imputed only for adults 25 years old or older.

Hot-deck imputation was used to impute missing values of household tenure and education attainment. Donors are randomly drawn from pools of hot-deck cells created using characteristics available for both donors and recipients. The search algorithm CHAID (Kass, 1980) was used to create the hot-deck cells using the variables found to be good predictors for the missing variable. A donor was then randomly drawn from the cell and the 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. Since household tenure is a household-level variable collected at the person level (a respondent could rent a room from an owner living in the same residence), and because adults are randomly selected within the household, person-level variables were not used in imputation for tenure. Table 8-4 shows the distribution of the imputed cases. When calculating the percentages the denominator for age group is the number of adults in the given age group, and for educational attainment the denominator is all adults.

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
ADLTCNT	Number of adults in the household
CHLDTOT	Number of children on the household
TEENTOT	Number of teens in the household
POVERTY	Poverty
P_GRAD	Percent college graduates in exchange
P_OWN	Percent home owners in the exchange
P_BLACK	Percent blacks in the exchange
P_HISP	Percent Hispanics in the exchange
Household Tenure	
ADCNT	Number of adults in the household
CHCNT	Number of children on the household
TNCNT	Number of teens in the household
P_GRAD	Percent college graduates in exchange
P_BLACK	Percent blacks in the exchange
P_HISP	Percent Hispanics in the exchange
P_65UP	Percent 65 and older in the exchange
P_OWN	Percent home owners in the exchange
POVERTY	Poverty

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

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

	Adult interviews	
	Count	%
Self-reported Education Attainment		
Under 25 years of age	NA	
Less than HS, 25 years of age or older	29	0.08
High School (or equivalent), 25 years of age or older	40	0.10
Some college, 25 years of age or older	43	0.11
BS and above, 25 years of age or older	70	0.18
Total	182	0.47
Self-reported Household Tenure		
Owner	269	0.64
Renter	146	0.35
Other Arrangement	19	0.05
Total	434	1.04

Source: UCLA Center for Health Policy Research, 2003 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 2003 California DOF Population Projections. The California DOF complies 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 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 respondents who could not identify with any of the five OMB race categories a sixth race category: "Some other race." 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." These procedures are described in the technical documentation of the MRDSF. The California DOF Projections used the 2000 MRDSF as the baseline for the time series; as a result, the DOF projections include only counts by the five OMB racial categories.

Following a procedure similar to the Census 2000, CHIS 2001 and 2003 respondents who could not identify themselves as any of the five OMB race categories could answer "some other race." 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 for 6,642 persons (12.2 percent) in CHIS 2003. After examining the procedures used by the Census to assign an OMB race, we determined that using the available variables in CHIS 2003, the assignment of OMB race could not be implemented as in Census 2000 because the number of CHIS cases in the geographic area by Hispanic origin[26] cells is not large enough to guarantee a good assignment. To reduce the number of records to be imputed, a combined race/ethnic variable (OMBSRREO) that grouped Latinos regardless of race into one level was proposed and approved by UCLA. This variable is all that is needed for raking. The levels of the variable OMBSRREO are given in Table 8-5.

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

Table 8-5. Levels of the OBM race-ethnic values (Variable OMBSRREO)

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

By creating a separate group for Latinos, a valid value of OMBSRREO was missing for only 128 persons (0.2%) who self-reported as non-Latino and "other race" alone[27] in CHIS 2003. 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 OMBSRRE for Latino eliminated the need to impute for 6,514 cases who reported Latino "other race" alone.

For continuity with the race and ethnicity variables created in CHIS 2001 (see Table 8-1), the same variables were created and imputed in CHIS 2003. We refer to these variables as the "regular" race and ethnicity variables. The OMB race-ethnicity variable OMBSRREO was created using the regular race and ethnicity variables (self-reported Hispanic SRH, self-reported White SRW, self-reported African American SRAA, self-reported American Indian Alaska Native SRAI, self-reported Asian SRAS, self-reported Hawaiian Native and Pacific Islander SRPI and self-reported other race SRO) 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 the 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 described below. One approach that was considered, but not adopted, was to impute the self-reported race and ethnicity of a respondent to any other sampled person within the household with missing values for these items. The reason this approach was not applied in CHIS 2001 or in CHIS 2003 was the realization that the method does not account for households with persons of more than one race and ethnicity.

[27] This includes records imputed as non-Hispanic "other" from the regular CHIS 2003 race imputation.

Instead a hot-deck imputation procedure was developed to deal with the diversity of race and ethnicity within households in a way the simpler assignment method does not. Before describing the hot-deck approach, some special 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 were there 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 percent of imputed interviews with missing self-reported race and/or ethnicity

	Imputed race*		Imputed ethnicity	
Type of interview	Count	%	Count	%
Adult	1,374	3.27	220	0.52
Child	415	4.87	46	0.54
Adolescent	274	6.83	52	1.30
Total	2,063	3.77	318	0.68

* At least one value of race was imputed.
Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

The hot-deck imputations were done separately by the 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 sampled (versus a household with an adult and an adolescent and/or a child). A household with only a sampled 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 sampled, 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 urbanicity considerations. 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 child are sampled. 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 child were sampled 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 analogous 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 and by type of extended interview

	Total		Extended interview type					
			Adult		Child		Adolescent	
	Count	%	Count	%	Count	%	Count	%
Self-reported race								
White alone	884	1.62	606	1.44	174	2.04	104	2.59
African American alone	41	0.08	31	0.07	8	0.09	2	0.05
Asian alone	92	0.17	52	0.12	12	0.14	28	0.70
American Indian/ Alaska Native alone	55	0.10	37	0.09	12	0.14	6	0.15
Pacific Islander alone	12	0.02	7	0.02	2	0.02	3	0.07
Other race alone	920	1.69	606	1.44	194	2.28	120	2.99
Two or more races	59	0.11	35	0.08	13	0.15	11	0.27
Total	2,063	3.78	1,374	3.27	415	4.87	274	6.83
Self Reported Ethnicity								
Latino	82	0.15	49	0.12	18	0.21	15	0.37
Non-Latino	236	0.43	171	0.41	28	0.33	37	0.92
Total	318	0.58	220	0.52	46	0.54	52	1.30
Completed interviews	54,580	100.00	42,044	100.00	8,526	100.00	4,010	100.00

Source: UCLA Center for Health Policy Research, 2003 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 by the Census Bureau. Since the OMB assignment is done using the imputed regular race variables at the time of assignment, all sampled persons have nonmissing races values (SRW, SRAA, SRAI, SRAS, SRPI, 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. 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 non-specified 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 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 "other" race are considered as single race respondents based on the OMB definition.

After the race/ethnicity assignments were made as described above, 128 persons (0.23 percent) remained with missing value 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 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 the regular race variables SRW, SRAA, SRAI, SRAS, and SRPI that were already imputed. 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. 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 race and ethnicity and type of extended interview (adult, child, and adolescent).

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

OMB Race-ethnicity (OMBSRREO)	Total		Extended interview type					
			Adult		Child		Adolescent	
	Count	%	Count	%	Count	%	Count	%
1. Latino	NA		NA		NA		NA	
2. Non-Latino White alone	90	0.16	75	0.18	7	0.08	8	0.20
3. Non-Latino African American alone	15	0.03	12	0.03	3	0.04	0	0.00
4. Non-Latino Asian alone	1	0.00	1	0.00	0	0.00	0	0.00
5. Non-Latino American Indian/ Alaska Native alone	13	0.02	10	0.02	2	0.02	1	0.02
6. Non-Latino Pacific Islander alone	0	0.00	0	0.00	0	0.00	0	0.00
7. Non-Latino two or more races	9	0.02	8	0.02	1	0.01	0	0.00
Total	128	0.23	106	0.25	13	0.15	9	0.22
Completed interviews	54,580		42,044		8,526		4,010	

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

8.4.3 Self-Reported Asian Ethnic Group

In CHIS 2003, the person weights were raked using a dimension defined for the Asian groups (Dimension 7). Although the total number of persons who self-reported Asian alone and in combination with another race(s) was controlled in CHIS 2001 through raking, subsequent analysis of the CHIS 2001 data showed differences from population totals of some Asian ethnic groups (e.g., Chinese and Filipino). For other Asian ethnic groups such as Korean and Vietnamese, separate weights or list-sample weights were created and controlled through raking dimensions in CHIS 2001. Since there is only one weight for the combined RDD and supplemental list samples in CHIS 2003, 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 group (Variable 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

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

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. 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 Asian group and type of extended interview (adult, child, and adolescent).

Table 8-11. Counts and percentage of imputed interviews with missing self-reported race and/or ethnicity by self-reported OMB Asian group and type of extended interview

OMB Asian group (OMBSRASO)	Total		Extended interview type: Adult		Child		Adolescent	
	Count	%	Count	%	Count	%	Count	%
1. Non-Latino Chinese alone	25	0.05	20	0.05	5	0.06	0	0.00
2. Non-Latino Korean alone	10	0.02	9	0.02	0	0.00	1	0.02
3. Non-Latino Filipino alone	10	0.02	8	0.02	2	0.02	0	0.00
4. Non-Latino Vietnamese alone	9	0.02	6	0.01	1	0.01	2	0.05
5. Other	79	0.14	46	0.11	17	0.20	16	0.40
Total	133	0.24	89	0.21	25	0.29	19	0.47
Completed interviews	54,580		42,044		8,526		4,010	

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

8.5 Self-Reported County and Self-Reported Stratum

In CHIS 2003, the location variables such as self-reported county of residence, Los Angeles SPA, city in Alameda County (Hayward, Oakland if respondent lives in Alameda County), Census tract, and the assignment of self-reported stratum were assigned after geocoding the geographic information collected during the interview (address of respondent, nearest street intersection, self-reported county) or attached to the sample telephone number (the mailing address or Zip Code covered by the telephone exchange). The 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, 2003 California Health Interview Survey.

The derived location variables SRSTRATA (self-reported stratum), SRCOUNTY (self-reported county), SR_LASPA (self-reported Los Angeles SPA), SR_ALCITY (self-reported city in Alameda County) are household-level variables that were assigned to all adult, child and adolescent records within the 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), SR_ALCITY (self-reported city in Alameda County) were used to create the cells for raking dimension 4 defined for Los Angeles County and Alameda County.

Table 8-13 shows the distribution of adult respondents by self-reported stratum compared with that at the time of sampling. 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. Net sample by sampling stratum and self-reported stratum

Strata	Sampling	Self-Reported	Net Agreement Ratio
Los Angeles	13,387	13,406	1.00
San Diego	2,975	2,988	1.00
Orange	2,898	2,836	1.02
Santa Clara	1,742	1,811	0.96
San Bernardino	1,672	1,676	1.00
Riverside	1,581	1,590	0.99
Alameda	6,087	5,972	1.02
Hayward	2,143	1,057	2.03
Oakland	2,489	2,337	1.07
Sacramento	1,344	1,342	1.00
Contra Costa	1,070	1,176	0.91
Fresno	870	873	1.00
San Francisco	1,068	1,051	1.02
Ventura	803	821	0.98
San Mateo	773	757	1.02
Kern	725	738	0.98
San Joaquin	697	701	0.99
Sonoma	642	655	0.98
Stanislaus	733	708	1.04
Santa Barbara	670	659	1.02
Solano	690	677	1.02
Tulare	799	810	0.99
Santa Cruz	647	609	1.06
Marin	646	647	1.00
San Luis Obispo	632	638	0.99
Placer	660	669	0.99
Merced	730	754	0.97
Butte	715	719	0.99
Shasta	630	671	0.94
Yolo	675	674	1.00
El Dorado	652	654	1.00
Imperial	738	736	1.00
Napa	635	650	0.98
Kings	764	759	1.01
Madera	684	676	1.01
Monterey, San Benito	686	712	0.96
Del Norte, Humboldt	664	660	1.01
Lassen, Modoc, Siskiyou, Trinity	521	529	0.98
Lake, Mendocino	512	498	1.03
Colusa, Glen, Tehama	556	517	1.08
Sutter, Yuba	617	601	1.03
Plumas, Nevada, Sierra	494	477	1.04
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	496	483	1.03

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Srvey

9. VARIANCE ESTIMATION

This chapter describes the methods and results of computing sampling errors for the CHIS 2003 RDD and surname list samples. 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 CHIS 2003. The remainder of the chapter describes the methodology for producing estimates of sampling variability from the survey. Section 9.1 discusses design effects and what they measure. Section 9.2 is a general review of the two main methods of computing sampling errors or variances of estimates from surveys with complex designs like CHIS 2003. 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 2003 estimates using commercially available software.

9.1 Design Effects

To evaluate the precision of sample estimates derived from a survey, sampling errors are computed from the data. 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. For instance, 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 of the estimated proportions.

Inferences of this nature require an estimate of the precision or sampling error of the characteristic being investigated. There are 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 for an estimate from a survey can be estimated as

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

We will refer to the estimated design effect as *DEFF*. At the analysis stage, the *DEFF* is useful because most statistical analysis software, such as SAS and SPSS, assume the data are from a simple random sample when computing sampling errors of estimates. The *DEFF* can, in some circumstances, indicate how appropriate this assumption is, and can be used to adjust these sampling errors of the estimates to produce ones that are closer to the actual sampling errors (Skinner, Holt, and Smith 1989).

The design effect for a proportion is particularly simple 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

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

where p is the estimated proportion, $v(p)_{SRS}$ is the estimated simple random sample variance $v(p)_{SRS} = \frac{p(1-p)}{n}$, and $v(p)_{COMPLEX}$ is the variance calculated appropriately from the survey.

In most surveys, design effects are larger than one. In CHIS 2003, design effects are greater than one mainly because the cases have different estimation weights (Kish 1992). As will be seen shortly, some design effects from the survey are considerably greater than one for statewide estimates.

Design effects are of primary interest to users of the data. They reveal that the complex sample design and estimation procedures used in the survey result in design effects that are greater than what would be found in a simple random sample. A simple random sample design was not considered for

CHIS 2003 because it would not have achieved the sample sizes for the specific domains of interest, in particular at the county/stratum level. The design effects calculated from the CHIS 2003 data indicate that the design and estimation procedures used in the survey need to be taken into account in the analysis of the data.

In CHIS 2003 and most large-scale surveys, a large number of data items or variables are collected. Each 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 is used to represent the design effects for similar items from the survey, as described in Wolter (1985).

The first panel in Tables 9-1 to 9-3 gives the average, median, maximum, and minimum *DEFF*s computed for a combination of categorical and continuous items. The second panel is the average *DEFT* for the same items. The *DEFT* is the square root of the design effect, so it is similar to the *DEFF* but on the scale of the standard error of the estimate rather than the variance.

Table 9-1 shows the *DEFF*s and *DEFT*s for 37 items selected from the adult interview by the county or stratum reported in the adult interview. Tables 9-2 and 9-3 present the corresponding *DEFF*s and *DEFT*s for 23 items from the child interviews and 25 items from the adolescent interview, respectively. Separate tables for the adult, child, and adolescent categorical variable estimates and the continuous variable estimates are given in Appendix C Table C-1, C-2, and C-3.

The *DEFT* is a more convenient measure than the *DEFF* because it can be used directly when computing confidence intervals for the estimates. In contrast, to use the *DEFF* when computing confidence intervals, the square root must be computed before it can be used. 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 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. By taking the square root of the *DEFF* and averaging these values, the variability is somewhat reduced. For example, in Table 9-1, the average *DEFF* for adults in Alameda is 2.46, while the maximum is 3.79 and the minimum is 1.64. This value is unusually large given the other values in the table. The average *DEFT* for Alameda is 1.57, which is also large, but not as different from the values for the other counties.

Before reviewing the tables in more detail, it is important to discuss the most important factors that result in design effects larger than one in CHIS 2003. These factors are:

- Oversampling. 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 differential sampling by listed and mailable status and refusal subsampling, but this had a relatively minor effect on the design effects.

- Within-Household Subsampling. Only one adult and one child or adolescent were 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.

- 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.

Table 9-1 shows the average *DEFT*s for estimates of adult items are between 1.08 to 1.25 in most counties. This implies that for most counties the standard error of the estimate is about 8 to 25 percent greater than the expected standard error of a simple random sample. The average *DEFT* is very consistent by county. The only exception is for Alameda County. The reason for the larger *DEFT* for Alameda County is that the cities of Hayward and Oakland were sampled at a much higher rate than the remainder of the county. When the data from the cities and the remainder of the county are combined, the cases have very different sampling rates and this increases the *DEFT*. The average *DEFT* for the state estimate is 1.28. This is slightly larger than the county-level *DEFT*s as expected because counties were not sampled proportional to their population. See Report 1: Sample Design for more details on the sampling.

Table 9-1. Average design effect (*DEFF*) and square root design effect (*DEFT*) for estimates from the adult interview*

Country/strata	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
State Total	1.69	1.75	3.01	0.55	1.28
Los Angeles	1.35	1.32	2.15	0.82	1.15
San Diego	1.55	1.56	2.24	0.89	1.24
Orange	1.58	1.52	2.37	0.94	1.25
Santa Clara	1.28	1.22	1.79	0.72	1.12
San Bernardino	1.36	1.31	2.13	0.85	1.16
Riverside	1.34	1.32	1.80	0.93	1.15
Alameda	2.46	2.38	3.79	1.64	1.56
Sacramento	1.41	1.37	2.43	0.74	1.18
Contra Costa	1.33	1.30	2.42	0.63	1.14
Fresno	1.35	1.35	2.15	0.67	1.15
San Francisco	1.27	1.27	1.87	0.63	1.12
Ventura	1.51	1.56	2.92	0.45	1.21
San Mateo	1.34	1.35	2.19	0.62	1.14
Kern	1.35	1.27	2.03	0.72	1.15
San Joaquin	1.35	1.35	1.92	0.49	1.15
Sonoma	1.30	1.31	1.88	0.76	1.13
Stanislaus	1.29	1.26	2.15	0.69	1.13
Santa Barbara	1.40	1.41	2.53	0.66	1.17
Solano	1.46	1.50	2.54	0.73	1.20
Tulare	1.23	1.23	1.87	0.78	1.11
Santa Cruz	1.32	1.29	1.94	0.75	1.14
Marin	1.30	1.34	2.25	0.53	1.13
San Luis Obispo	1.23	1.21	1.80	0.74	1.10
Placer	1.37	1.29	2.33	0.80	1.16
Merced	1.40	1.38	2.05	0.82	1.17
Butte	1.24	1.21	1.67	0.75	1.11
Shasta	1.27	1.28	2.10	0.77	1.12
Yolo	1.33	1.32	1.89	0.64	1.14
El Dorado	1.21	1.20	1.84	0.49	1.09
Imperial	1.35	1.30	2.08	0.74	1.15
Napa	1.38	1.31	3.89	0.26	1.15
Kings	1.30	1.28	1.99	0.36	1.13
Madera	1.29	1.33	1.87	0.67	1.13
Monterey, San Benito	1.23	1.24	2.04	0.76	1.10
Del Norte, Humboldt	1.19	1.18	1.75	0.70	1.08
Lassen, Modoc, Siskiyou, Trinity	1.22	1.24	1.94	0.63	1.10
Lake, Mendocino	1.26	1.23	1.79	0.76	1.12
Colusa, Glen, Tehama	1.18	1.11	2.83	0.51	1.08
Sutter, Yuba	1.21	1.20	2.05	0.50	1.09
Plumas, Nevada, Sierra	1.38	1.34	2.46	0.63	1.16
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.23	1.22	1.91	0.77	1.11

* Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.
Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

Table 9-2 shows the average *DEFT* for estimates from the child interview in each county. 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 (i.e., the number of children per household is more variable than the number of adults per household). The average *DEFT* at the state level is 1.37. The average *DEFT*s for the counties are between 1.03 and 1.30; that is, the standard errors of these estimates are between 3 and 30 percent greater than expected from a simple random sample. The only exception is again Alameda County for the same reason as noted above for the adult interview items.

Table 9-2. Average design effect (*DEFF*) and square root design effect (*DEFT*) for estimates from the child interview*

Country/strata	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
State Total	1.90	1.97	2.34	1.35	1.37
Los Angeles	1.61	1.57	2.19	1.19	1.26
San Diego	1.60	1.59	2.26	1.18	1.26
Orange	1.71	1.69	2.48	0.87	1.30
Santa Clara	1.56	1.44	2.61	0.63	1.24
San Bernardino	1.65	1.50	2.93	1.13	1.27
Riverside	1.38	1.33	1.96	0.94	1.17
Alameda	2.56	2.51	3.73	1.49	1.59
Sacramento	1.27	1.30	1.81	0.49	1.12
Contra Costa	1.38	1.37	2.41	0.83	1.16
Fresno	1.44	1.46	1.95	0.79	1.19
San Francisco	1.43	1.38	2.38	0.64	1.18
Ventura	1.39	1.39	2.68	0.22	1.15
San Mateo	1.31	1.23	2.97	0.45	1.12
Kern	1.35	1.33	2.25	0.75	1.15
San Joaquin	1.43	1.40	2.34	0.68	1.19
Sonoma	1.17	1.29	1.53	0.54	1.07
Stanislaus	1.42	1.43	1.98	0.86	1.18
Santa Barbara	1.46	1.44	2.22	0.63	1.19
Solano	1.51	1.39	5.19	0.49	1.19
Tulare	1.29	1.25	1.72	0.86	1.13
Santa Cruz	1.40	1.39	2.03	0.91	1.18
Marin	1.19	1.19	1.90	0.61	1.08
San Luis Obispo	1.33	1.36	1.88	0.87	1.15
Placer	1.39	1.33	3.31	0.89	1.17
Merced	1.61	1.55	2.58	1.01	1.26
Butte	1.28	1.28	2.04	0.21	1.12
Shasta	1.16	1.12	1.79	0.73	1.07
Yolo	1.25	1.29	1.71	0.69	1.11

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

County/strata	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
El Dorado	1.09	1.08	1.72	0.45	1.03
Imperial	1.23	1.19	2.09	0.71	1.10
Napa	1.45	1.16	3.20	0.67	1.18
Kings	1.36	1.39	1.91	0.12	1.15
Madera	1.30	1.36	1.83	0.58	1.13
Monterey, San Benito	1.37	1.35	2.61	1.00	1.16
Del Norte, Humboldt	1.21	1.08	2.53	0.37	1.08
Lassen, Modoc, Siskiyou, Trinity	1.55	1.52	2.31	0.45	1.23
Lake, Mendocino	1.16	1.18	1.55	0.59	1.07
Colusa, Glen, Tehama	1.26	1.23	2.05	0.44	1.11
Sutter, Yuba	1.46	1.43	2.16	0.92	1.20
Plumas, Nevada, Sierra	1.61	1.63	2.33	0.95	1.26
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.21	1.25	1.73	0.52	1.09

* Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.
Source: UCLA Center for Health Policy Research, 2003 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 so similar to that of children we should expect a close correspondence between the two. The average *DEFT* for the state estimates is 1.40. For most of the strata, the average *DEFT*s are between 1.01 and 1.40.

Table 9-3. Average design effect (*DEFF*) and square root design effect (*DEFT*) for estimates from the adolescent interview*

County/strata	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
State Total	1.98	1.92	3.50	1.16	1.40
Los Angeles	1.86	1.86	2.58	1.27	1.36
San Diego	1.63	1.68	2.47	0.58	1.27
Orange	1.88	1.86	2.88	1.16	1.36
Santa Clara	1.74	1.83	3.61	0.30	1.29
San Bernardino	1.31	1.25	2.35	0.81	1.14
Riverside	1.47	1.26	4.07	0.71	1.19
Alameda	2.01	2.07	3.16	0.42	1.39
Sacramento	1.23	1.21	2.25	0.50	1.09
Contra Costa	1.33	1.34	1.87	0.63	1.14
Fresno	1.47	1.12	2.95	0.87	1.19
San Francisco	1.17	1.08	1.98	0.56	1.07
Ventura	1.39	1.39	2.78	0.45	1.16

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

County/strata	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average
San Mateo	1.57	1.71	2.76	0.60	1.23
Kern	1.28	1.30	1.72	0.77	1.13
San Joaquin	1.10	1.04	1.87	0.50	1.04
Sonoma	1.23	1.27	2.58	0.57	1.09
Stanislaus	1.71	1.80	3.26	0.37	1.28
Santa Barbara	1.21	1.31	1.63	0.41	1.09
Solano	1.42	1.32	4.17	0.72	1.17
Tulare	1.37	1.43	2.09	0.40	1.16
Santa Cruz	1.20	1.18	1.96	0.57	1.08
Marin	1.04	0.97	2.06	0.59	1.01
San Luis Obispo	1.40	1.33	3.17	0.60	1.16
Placer	1.14	1.22	1.77	0.35	1.06
Merced	1.47	1.48	2.51	0.84	1.20
Butte	1.30	1.22	2.15	0.82	1.13
Shasta	1.29	1.20	2.25	0.80	1.13
Yolo	1.16	1.24	1.62	0.40	1.07
El Dorado	1.18	1.20	2.15	0.52	1.07
Imperial	1.37	1.27	2.92	0.00	1.13
Napa	1.57	1.64	2.51	0.39	1.23
Kings	2.05	1.92	4.06	0.43	1.40
Madera	1.59	1.46	3.22	0.83	1.25
Monterey, San Benito	1.29	1.36	2.36	0.64	1.12
Del Norte, Humboldt	1.10	1.06	1.82	0.57	1.04
Lassen, Modoc, Siskiyou, Trinity	1.53	1.41	4.56	0.79	1.22
Lake, Mendocino	1.20	1.21	1.68	0.88	1.09
Colusa, Glen, Tehama	1.45	1.35	2.23	0.80	1.19
Sutter, Yuba	1.69	1.69	3.74	0.38	1.26
Plumas, Nevada, Sierra	1.29	1.18	2.10	0.85	1.13
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.12	1.08	1.88	0.82	1.05

* Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.0.82.
Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.0.85.

9.2 Methods for Variance Estimation

Variance estimation procedures have been developed to account for the sample design employed in a complex survey. Using these procedures, factors such as the selection of sample clusters in 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 for estimating variances from a complex survey are replication methods and the Taylor series approximation method. Wolter

(1985) is a useful reference on the theory and applications of these methods. Shao (1996) is a more recent review paper that compares the 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 estimates of variance for sample estimates of interest. Estimates of variance of an arbitrary parameter of interest, θ ,take the following form

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

where

$\hat{\theta}$	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, and
c	is a constant that depends on the replication method.

In the next section, the specific form of equation (1) used in CHIS 2003 is presented.

The other widely used method for estimating variances in complex surveys is based on the Taylor series approximation. A 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. The Taylor series method relies on the simplicity associated with estimating the variance for a linear statistic even with a complex sample design. In most complex designs, the variance can be estimated by using the

variance between primary sampling units (PSUs) and a with-replacement design (Wolter 1985). In this formulation, the strata and PSUs must be defined, similar to the variance estimation strata and units discussed above.

9.3 Design of Replicates

In CHIS 2003, a paired unit jackknife which is a form of the jackknife replication method (JK2[28]) was selected for computing variances from the survey data. The following sections provide the 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 2003 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 2003 estimates affect the sampling errors of the estimates produced from the survey. The replicate weights prepared for the survey reflect all such aspects of weighting. 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 then in specialized situations.

Adjacent pairs of sampled telephone numbers were treated as having been sampled from the same stratum. The details of the assignment are given below. The same approach was used for another RDD study, the 1993 National Household Education Survey (Brick et al., 1997). The JK2 approach treats each pair of sampled telephone numbers as an implicit stratum, where each such stratum is defined by the sort order used in the sample selection of telephone numbers. In the JK2 method, the constant, c, in equation (1) is equal to 1.

[28] 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 strata, so this really involves specifying 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 2003, we elected to create 80 variance estimation strata, even though many more could have been created. 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 80^{th} pair to stratum 80, the 81^{st} pair to stratum 1, the 82^{nd} pair to stratum 2, etc.). As a result, each variance stratum had approximately the same number of telephone numbers. The same process was followed for each sampling stratum.

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

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

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

Many standard statistical software packages assume a simple random sample when computing estimates of variance. As a result, estimates of variance from these packages can seriously understate the true variability of the survey estimates. 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 2003 using some of these programs beginning with the program used to compute the sampling errors in this report, WesVar.

WesVar™ Version 4.1 (Westat, 2000) is a software package developed and distributed by Westat. WesVar™ uses replication methods to compute variance estimates. Through the use of replicates, adjustments made during weighting (nonresponse and raking) can be taken into account by making the same adjustments to each replicate separately. Replication is computer intensive, but powerful personal computers have largely eliminated this as an issue for all but the largest data sets and most complicated analyses. Although replication can be used for most estimates, replication techniques are not necessarily appropriate for all sample statistics of interest. Special care is needed when trying to estimate the median, quartiles, or other quantiles. WesVar™ computes sampling errors of quantiles using an approximation method that has relatively good statistical properties.

WesVar™ is an interactive program with a graphical interface that makes it simple to specify the estimates for which sampling errors are needed. 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. The information about the design is incorporated into the replicate weights when the data file is created. Descriptive statistics of analysis variables can be produce through "table requests" in WesVar. 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 2003 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 will properly account for the sample design and estimation methods because these features are accounted for in the replicate weights.

SUDAAN® (Software for the Statistical Analysis of Correlated Data) (Research Triangle Institute, 2001) is a package developed by Research Triangle Institute to analyze data from complex sample surveys. Like WesVar™, SUDAAN® computes standard errors of the estimates taking into account the survey design. SUDAAN® and WesVar™ produce the same point estimates. The difference is in the method used to compute the variances. SUDAAN® uses a first-order Taylor series approximation, although some replication methods are included in later versions.[29] When the Taylor series approximations are used, SUDAAN® does not fully take into account complex weighting schemes such as nonresponse adjustments or raking. Medians and quantiles cannot be computed directly using either the Taylor series method or the replication method, but the same type of approximation is used in WesVar and SUDAAN® to compute medians.

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. SUDAAN® also contains procedures for computing other analytic statistics, such as those associated with linear and regression models.

SAS® has also introduced new procedures to analyze survey data. SAS® Version 8 (SAS Institute, 2002) has two procedures for analyzing survey data: PROC SURVEYMEANS and PROC SURVEYREG. Both use the Taylor series linearization approach to estimate standard errors. SURVEYMEANS produces estimates of means, proportions, and totals, while SURVEYREG fits linear regression models (logistic regression is not available). No design effects are estimated with either PROC. Estimates of differences or other linear combinations are not available in SURVEYMEANS. These procedures are relatively new in SAS and do not contain as many features as most of the other packages. At the current time, the SAS procedures are the most limited of all the packages we discuss.

Another software package that can be used to analyze survey data is STATA® (Stata Corporation, 2001). 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.

[29] To use the replication methods in SUDAAN with the CHIS data you must specify the following parameters in each run: DESIGN=JACKKNIFE; WEIGHT (fullsample weight); JACKWGTS (replicate weights) /ADJJACK=1.

STATA® has a family of *svy*- commands to analyze data from sample surveys. Some STATA® commands used to analyze survey data are **svytotal**, **svyprop, svytab, svymean, svylc,** and **svylogit**. These are used to estimate 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 **svytab.** The command **svymean** can be used to produce the *DEFF*s for proportions by coding the analytical variable with values 0 and 1. To estimate totals using **svytotal**, a variable ONE must be created with a value of 1 for all the records.

STATA® uses the Taylor series method of variance estimation. Like the other Taylor series software it cannot account for nonresponse or raking. 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. STATA® supports more analytic methods than any of the other packages.

All of the programs that use the Taylor series approximation require auxiliary variables that provide information about the sample design. To support analysis with the Taylor series method, two variables have been defined and included in the data files. The two variables reflect the original sample design and are required in all the Taylor series software packages. The variables are:

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

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

The same variables, TSVAR and TSUNIT, can be used for SUDAAN®, SAS®, and STATA®.

REFERENCES

Anderson, J.E., Nelson, D.E., and Wilson, R.W. (1998). Telephone coverage and measurement of health risk indicators: Data from the National Health Interview Survey. *American Journal of Public Health*, 88, 1392-1395.

Blumberg, S., Luke, J., and Cynamon, M. (2004). Has cord-cutting cut into random-digit-dialed health surveys? The prevalence and impact of wireless substitution. *Proceedings of the Eight Conference on Health Survey Research Methods*, Atlanta, GA.

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

Brick, J. M., E. Tubbs, M. Collins, and M. J. Nolin. (1997). Unit and item response, weighting and imputation procedures in the 1993 National Household Education Survey, Working Paper 97-05. U.S. Department of Education, Office of Educational Research and Improvement.

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

Brick, J.M., Judkins, D., Montaquila, J., and Morganstein, D. (2002). Two-phase list-assisted RDD sampling. *Journal of Official Statistics*, Vol. 18, No. 2, 203-215.

Brick, J.M., Montaquila, J., and Scheuren, F.J. (2002). Estimating residency rates for undetermined telephone numbers, *Public Opinion Quarterly*, Vol. 66, 18-39.

Brick, J.M., Waksberg, J., and Keeter, S. (1996). Using data on interruptions in telephone service as coverage adjustments, *Survey Methodology*, 22(2), 185-197.

Brick, J.M., Montaquila, J., Collins-Hagedorn, M., Brock-Roth, S., and Chapman, C. (forthcoming). Implications for RDD Design From An Incentive Experiment, *Journal of Official Statistics*.

Council of American Survey Research Organizations. (1982). *On the definition of response rates* (special report of the CASRO task force on completion rates, Lester R. Frankel, Chair). Port Jefferson, NY: Author.

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

Ferraro, D., and Brick J.M. (2001). Weighting for nontelephone households in RDD surveys. *Proceedings of the Survey Research Methods Section of the American Statistical Association*.

Ford, B. M. (1983), An Overview of Hot-Deck Procedures. Incomplete Data in Sample Surveys, Vol. 2. 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.

Holt, D., and Smith, T. (1979). Poststratification. *Journal of Royal Statistical Society* A (142), 33-46.

Kalton, G., and Flores Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics* Vol. 19, No. 2, 81-97.

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

Keeter, S. (1995). Estimating telephone noncoverage bias from a phone survey. *Public Opinion Quarterly*, 59, 196-217.

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

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

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," Vol. 62, No. 210, Thursday, October 30, 1997.

Oh, H.L., and Scheuren, F.J. (1983). Weighting adjustments for unit nonresponse. In *Incomplete Data in Sample Surveys* Volume 2, W.G. Madow, I. Olkin, and D.B. Rubin (Eds.). Academic Press, 143-184.

Research Triangle Institute (2001). *SUDAAN User's Manual*, *Release 8.0*, Research Triangle Institute, Research Triangle Park, NC.

Rizzo, Lou, J. Michael Brick and Inho Park. (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 (2002). *SAS OnlineDoc® Version 8,* SAS Institute, Inc., Cary, NC.

Shao, J. (1996). Resampling Methods in Sample Surveys, (with Discussion). *Statistics*, 27, 203-254.

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

State of California, Department of Finance, (2004). *Population Projections by Race/Ethnicity for California and Its Counties 2000-2050,* Sacramento California.

Stata Corporation (2001), *STATA User's Guide*, *Release 7*, Stata Corporation, College Station, TX.

Tucker, C., Brick, J.M., Meekins, B., and Morganstein, D. (2004). Household Telephone Service and Usage Patterns in the U.S. in 2004, *Proceedings of the Survey Methods Section of the American Statistical Association* [CD-ROM].

U.S. Census Bureau (2001). 2000 Summary File 1, California, prepared by the U.S. Census Bureau.

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, California, prepared by the U.S. Census Bureau.

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 2003 frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean and Vietnamese lists).

Sampling stratum	Description	RDD sampling frame			Korean surname list			Vietnamese surname list			Korean/Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
1.012	San Fernando Valley (High Minority)	157,100	4,000	39.16	1,920	24	28.24	217	1	27.13	1,484	33	23.56
1.013	San Gabriel Valley (High Minority)	346,200	6,300	54.87	10,855	112	36.43	5,733	47	38.22	5,138	104	25.56
1.014	Metro (High Minority)	365,300	10,500	34.87	7,176	87	23.30	506	1	42.17	4,260	101	18.21
1.015	West (High Minority)	65,400	1,700	38.47	402	4	23.65	201	3	18.27	246	3	61.50
1.017	East (High Minority)	90,000	2,900	31.09	1,823	16	22.79	326	3	21.73	1,129	23	19.47
1.018	South Bay (High Minority)	174,700	4,500	38.79	1,496	15	28.23	637	5	28.96	1,046	20	24.33
1.021	Antelope Valley (Low Minority)	209,700	5,000	41.95	200	3	25.00	180	2	22.50	182	2	36.40
1.022	San Fernando Valley (Low Minority)	1,552,100	17,700	87.74	4,649	50	43.45	2,403	18	51.13	3,537	70	30.76
1.023	San Gabriel Valley (Low Minority)	915,700	12,300	74.38	7,546	78	44.39	5,007	46	44.71	3,943	80	29.87
1.024	Metro (Low Minority)	667,600	8,700	76.78	4,084	40	42.10	1,766	16	42.05	2,515	57	28.58
1.025	West (Low Minority)	960,200	10,400	92.41	3,135	29	44.79	936	11	36.00	1,683	25	41.05

Table A-1. CHIS 2003 frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean and Vietnamese lists) (continued)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Vietnamese surname list			Korean/Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
1.026	South (Low Minority)	654,000	7,800	83.94	1,431	16	38.68	422	3	42.20	1,188	24	28.98
1.027	East (Low Minority)	761,900	8,900	85.59	2,494	23	48.90	837	6	46.50	1,558	35	31.16
1.028	South Bay (Low Minority)	1,043,500	13,500	77.33	3,659	34	43.05	2,381	22	42.52	2,428	51	29.61
2.01	San Diego (High Minority)	84,000	3,400	24.53	546	3	23.74	1,464	16	23.61	548	11	21.92
2.02	San Diego (Low Minority)	2,078,100	17,000	122.11	4,612	42	63.18	4,801	54	57.84	3,245	64	36.46
3.01	Orange (High Minority)	741,400	13,700	54.10	5,892	71	37.29	15,981	166	33.16	6,407	139	25.22
3.02	Orange (Low Minority)	1,767,400	10,300	171.35	6,245	72	64.38	5,870	53	67.47	4,595	91	39.96
4.01	Santa Clara (High Minority)	739,500	7,600	97.35	5,037	45	51.40	13,023	130	49.33	4,840	101	31.63
4.02	Santa Clara (Low Minority)	846,100	5,700	148.15	7,150	83	55.00	2,689	31	62.54	3,199	60	41.01
5	San Bernardino	1,197,200	10,300	116.38	2,466	21	52.47	1,931	17	50.82	1,871	41	30.67
6	Riverside	1,167,900	9,600	121.69	1,924	21	53.44	1,683	17	54.29	1,344	22	37.33
7.10711	Hayward* Hayward (High AfAm)	59,900	9,083	6.59	165	1	6.60	299	1	6.10	139	5	6.95
7.10712	Hayward* Hayward (Low AfAm)	7,600	638	11.89	16	1	4.00	40	1	20.00	15	-	15.00
7.10721	Hayward* Oakland (High AfAm)	109,100	17,419	6.27	637	6	5.44	392	3	5.44	353	4	6.30

Table A-1. CHIS 2003 frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean and Vietnamese lists) (continued)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Vietnamese surname list			Korean/Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
7.10722	Hayward*Oakland (Low AfAm)	39,700	3,566	11.11	287	2	11.96	228	1	10.36	173	1	13.31
7.10731	Hayward*Balance (High AfAm)	19,800	2,158	9.20	147	3	12.25	109	-	5.45	78	1	5.20
7.10732	Hayward*Balance (Low AfAm)	65,700	2,441	26.86	619	10	17.69	384	4	25.60	285	4	25.91
7.2072	Oakland	448,800	20,405	22.05	2,696	18	18.34	2,510	19	16.96	1,798	41	11.68
7.3072	Balance Alameda*Oakland	300,300	47,665	6.30	2,760	19	6.13	1,124	3	6.18	1,385	19	5.27
7.3073	Balance Alameda*Balance Alameda	433,500	9,925	43.65	3,072	34	28.18	1,233	11	31.62	1,341	25	30.48
8	Sacramento	1,061,400	8,700	121.93	2,853	25	63.40	3,825	30	67.11	2,341	35	41.07
9	Contra Costa	890,000	7,100	125.18	2,571	21	69.49	1,457	16	52.04	1,606	32	41.18
10	Fresno	650,600	5,800	112.26	1,431	11	49.34	939	8	52.17	801	15	40.05
11	San Francisco	1,068,400	12,600	84.61	12,888	140	45.86	5,901	65	42.76	6,188	101	39.92
12	Ventura	596,500	5,700	104.57	1,059	13	50.43	782	12	41.16	708	8	54.46
13	San Mateo	790,200	6,500	121.47	3,952	43	56.46	1,333	14	45.97	1,943	35	41.34
14	Kern	507,900	4,100	123.94	516	6	46.91	368	2	61.33	359	9	29.92
15	San Joaquin	391,800	4,100	95.44	895	12	40.68	1,273	16	45.46	625	9	56.82
16	Sonoma	434,300	3,900	111.28	627	8	52.25	522	4	65.25	404	4	57.71
17	Stanislaus	310,000	4,000	77.49	432	5	43.20	374	2	46.75	263	5	37.57
18	Santa Barbara	379,700	4,200	90.38	596	7	39.73	352	4	44.00	316	3	79.00

Table A-1. CHIS 2003 frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean and Vietnamese lists) (continued)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Vietnamese surname list			Korean/Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
19	Solano	283,100	3,500	80.85	419	3	41.90	403	1	100.75	363	8	33.00
20	Tulare	282,700	4,800	58.94	212	2	42.40	134	1	26.80	115	2	19.17
21	Santa Cruz	257,600	4,000	64.35	374	1	53.43	180	-	90.00	225	5	37.50
22	Marin	325,300	5,200	62.50	518	7	34.53	355	4	44.38	339	4	48.43
23	San Luis Obispo	233,200	3,400	68.51	254	5	42.33	219	2	73.00	189	3	27.00
24	Placer	262,700	3,900	67.38	330	5	30.00	241	2	40.17	242	7	24.20
25	Merced	119,100	3,900	30.54	189	1	27.00	120	-	20.00	138	4	19.71
26	Butte	159,100	3,200	49.75	242	3	40.33	209	4	19.00	148	1	29.60
27	Shasta	139,300	3,400	40.92	130	2	32.50	98	-	98.00	107	3	26.75
28	Yolo	138,500	3,300	42.07	877	10	24.36	529	4	33.06	508	12	18.81
29	El Dorado	144,200	4,000	36.02	177	1	44.25	107	-	53.50	131	3	21.83
30	Imperial	84,300	4,100	20.57	132	1	16.50	33	-	11.00	81	1	20.25
31	Napa	114,400	3,900	29.33	98	1	49.00	71	2	35.50	79	1	15.80
32	Kings	70,400	4,000	17.60	71	2	17.75	50	2	10.00	52	-	13.00
33	Madera	87,300	4,200	20.79	66	2	16.50	36	-	18.00	43	-	14.33
34	Monterey, San Benito	388,200	5,000	77.64	664	4	60.36	369	3	46.13	441	7	31.50
35	Del Norte, Humboldt	148,400	4,100	36.18	138	1	34.50	101	1	50.50	96	1	19.20
36	Lassen, Modoc, Siskiyou, Trinity	95,100	3,500	27.16	81	-	27.00	57	1	28.50	52	2	17.33

Table A-1. CHIS 2003 frame sizes[1], sample sizes[2], and base weights by sampling stratum and sampling frame (RDD, Korean and Vietnamese lists) (continued)

Sampling stratum	Description	RDD sampling frame			Korean surname list			Vietnamese surname list			Korean/Vietnamese surname list		
		Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight	Frame	Sample	Weight
37	Lake, Mendocino	129,900	3,200	40.56	164	2	54.67	92	-	30.67	67	2	33.50
38	Colusa, Glen, Tehama	83,500	2,900	28.80	60	-	20.00	62	-	31.00	43	-	43.00
39	Sutter, Yuba	100,700	3,100	32.46	202	1	28.86	143	-	47.67	88	2	22.00
40	Plumas, Nevada, Sierra	135,900	3,100	43.85	136	2	27.20	93	1	31.00	117	-	29.25
41	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	226,000	3,700	61.08	177	-	44.25	129	-	64.50	108	1	54.00

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

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

Source: UCLA Center for Health Policy Research, 2003 California Health Interview Survey.

Appendix B

Table B–1. Household weighting for the combined RDD and surname list samples by stratum

		All strata	Los Angeles	San Diego	Orange	Santa Clara
1.	Base weight					
	1.1 Sample size	463,025	115,544	20,590	24,592	13,750
	1.2 Sum of weight	29,159,693	7,963,855	2,162,100	2,508,800	1,585,600
	1.3 Coefficient of variation (CV)	68.09	30.81	35.06	57.46	25.71
2.	Adjusting for mailable subsampling					
	2.1 Sample Size before mailable subsampling					
	a. Mail	190,893	47,522	8,869	9,457	5,444
	b. Non-mail	102,322	27,614	4,274	5,734	2,749
	c. Nonresidential	170,264	40,862	7,447	9,401	5,557
	2.2 Sample Size after mailable subsampling					
	a. Mail	150,901	43,746	8,291	8,795	4,933
	b. Non-mail	56,157	16,916	2,946	3,865	1,388
	c. Nonresidential	125,383	30,759	5,581	7,046	4,173
	2.3 Sample size	332,441	91,421	16,818	19,706	10,494
	2.4 Sum of weights after adjustment					
	a. Mail	11,784,286	3,240,289	903,852	951,299	596,865
	b. Non-mail	5,562,516	1,647,962	425,879	544,793	244,361
	c. Nonresidential	11,812,891	3,075,603	832,369	1,012,708	744,375
	2.5 Sum of weights	29,159,693	7,963,855	2,162,100	2,508,800	1,585,600
	2.6 Coefficient of variation (CV)	69.44	46.35	46.40	65.82	47.18
	2.7 Observed subsampling rates					
	a. Mail subsampling rate	0.791	0.921	0.935	0.930	0.906
	b. Non-mail subsampling rate	0.549	0.613	0.689	0.674	0.505
	c. Nonresidential subsampling rate	0.736	0.753	0.749	0.749	0.751

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	All strata	Los Angeles	San Diego	Orange	Santa Clara
3. Refusal Subsampling Adjustment					
3.1 Sum of weights before adjustment					
a. Never refused	23,758,933	6,441,770	1,719,893	2,033,718	1,339,716
b. Refused and was selected for refusal conversion	3,696,375	1,038,409	287,821	318,611	177,603
c. Refused and was not selected for refusal conversion	1,704,385	483,676	154,385	156,471	68,281
3.2 Sum of weights after adjustment					
a. Never refused	23,758,933	6,441,770	1,719,893	2,033,718	1,339,716
b. Refused and was selected for refusal c. conversion	5,400,760	1,522,085	442,207	475,082	245,884
c. Refused and was not selected for refusal conversion	0	0	0	0	0
3.3 Sum of weights	29,159,693	7,963,855	2,162,100	2,508,800	1,585,600
3.4 Sample size	310,233	84,728	15,455	18,344	9,965
3.5 Coefficient of variation (CV)	70.07	45.80	51.88	68.58	39.36
4. Adjusting for unknown residential					
4.1 Sum of weights by residential status before adjustment					
a. Residential – respondent	6,183,984	1,562,438	483,965	483,262	298,225
b. Residential – nonrespondent	3,891,573	1,232,637	294,297	327,051	169,445
c. Unknown residential status-NA,NM	2,703,329	743,064	210,949	240,549	149,967
d. Nonresidential	16,266,689	4,391,342	1,161,640	1,446,558	963,623
4.2 Sum of weights before adjustment					
a. Residential – respondent	6,183,984	1,562,438	483,965	483,262	298,225
b. Residential – nonrespondent	3,891,573	1,232,637	294,297	327,051	169,445
c. Estimated Residential NA, NM	990,423	264,629	73,912	81,966	54,867
4.3 Sum of weights after adjustment					
a. Residential – respondent	6,183,984	1,562,438	483,965	483,262	298,225
b. Residential – nonrespondent	4,881,996	1,498,045	367,878	408,280	224,216
c. Unknown residential status-NA,NM	0	0	0	0	0
4.4 Sum of weights	11,065,980	3,060483	851,843	891,543	522,441
4.5 Sample size	103,796	30,204	5,553	5,861	3,192
4.6 Coefficient of variation (CV)	91.38	71.98	78.63	96.38	75.26

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	All strata	Los Angeles	San Diego	Orange	Santa Clara
5. Unknown minority screener adjustment					
5.1 Sum of weights before adjustment					
a. Screener minority or other strata	10,954,716	3,027,335	846,275	878,162	508,674
b. Non-screener minority	22,026	5,581	693	1,567	2,668
c. Unknown screener minority	89,238	27,568	4,875	11,814	11,099
5.2 Sum of weights after adjustment					
a. Screener minority or other strata	10,997,924	3,039,278	849,138	889,909	516,501
b. Non-screener minority	68,056	19,631	2,444	5,519	9,397
c. Unknown screener minority	0	0	0	0	0
6. Screener nonresponse adjustment					
6.1 Sum of weights before adjustment	11,065,980	3,060,483	851,843	891,543	522,441
a. Respondents	6,201,865	1,567,870	485,764	492,627	303,275
b. Nonrespondents	4,864,116	1,491,039	365,818	402,801	222,623
6.2 Sum of weights after adjustment	11,065,980	3,058,909	851,582	895,428	525,899
7. Multiple telephone adjustment					
7.1 Sum of weights before adjustment	11,065,980	3,060483	851,843	891,543	522,441
7.2 Sum of weights after adjustment	10,541,421	2,906,878	812,890	849,343	500,663
7.3 Adjustment factor	0.95	0.95	0.95	0.95	0.96
8. Household poststratification					
8.1 Number of completed screeners	66,657	17,770	3,701	3,674	2,042
8.2 Sum of weights before adjustment	10,541,421	2,906,878	812,890	849,343	500,663
8.3 Sum of weights after adjustment	11,502,870	3,133,774	994,677	935,287	565,863
8.4 Adjustment factor	1.09	1.08	1.22	1.10	1.13
8.5 Coefficient of variation (CV)	64.70	42.32	48.32	67.23	38.02
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment	11,502,870	3,133,774	994,677	935,287	565,863
a. Completed section-G	7,293,156	7,293,343	1,841,789	634,741	575,571
b. Did not complete section-G	4,209,714	4,209,527	1,291,985	359,936	359,716
9.2 Sum of weights after adjustment	11,502,870	3,133,774	994,677	935,287	565,863
9.3 Number of adults completed through section G	66,657	17,770	3,701	3,674	2,042
9.4 Coefficient of variation (CV)	66.90	100.23	97.69	117.32	84.45

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
1. Base weight					
1.1 Sample size	10,379	9,660	113,537	8,790	7,169
1.2 Sum of weight	1,197,200	1,167,900	1,484,400	1,061,400	890,000
1.3 Coefficient of variation (CV)	7.58	6.29	87.12	7.32	6.94
2. Adjusting for mailable subsampling					
2.1 Sample Size before mailable subsampling					
a. Mail	4,396	4,284	47,706	3,584	2,982
b. Non-mail	2,565	2,375	24,730	1,947	1,264
c. Nonresidential	3,418	3,001	41,101	3,259	2,923
2.2 Sample Size after mailable subsampling					
a. Mail	3,943	3,843	19,190	3,231	2,730
b. Non-mail	1,286	1,206	9,523	1,242	627
c. Nonresidential	2,560	2,253	28,329	2,446	2,192
2.3 Sample size	7,789	7,302	57,042	6,919	5,549
2.4 Sum of weights after adjustment					
a. Mail	514,385	527,365	588,376	415,713	370,005
b. Non-mail	228,319	223,164	312,805	216,648	115,633
c. Nonresidential	454,495	417,371	583,220	429,040	404,362
2.5 Sum of weights	1,197,200	1,167,900	1,484,400	1,061,400	890,000
2.6 Coefficient of variation (CV)	41.49	41.11	110.13	33.62	36.56
2.7 Observed subsampling rates					
a. Mail subsampling rate	0.897	0.897	0.402	0.902	0.915
b. Non-mail subsampling rate	0.501	0.508	0.385	0.638	0.496
c. Nonresidential subsampling rate	0.749	0.751	0.689	0.751	0.750

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
3. Refusal Subsampling Adjustment					
3.1 Sum of weights before adjustment					
a. Never refused	934,990	901,493	1,230,860	870,367	728,662
b. Refused and was selected for refusal conversion	196,075	194,704	147,993	131,021	115,120
c. Refused and was not selected for refusal conversion	66,135	71,703	105,547	60,012	46,219
3.2 Sum of weights after adjustment					
a. Never refused	934,990	901,493	1,230,860	870,367	728,662
b. Refused and was selected for refusal c. conversion	262,210	266,407	253,540	191,033	161,338
c. Refused and was not selected for refusal conversion	0	0	0	0	0
3.3 Sum of weights	1,197,200	1,167,900	1,484,400	1,061,400	890,000
3.4 Sample size	7,298	6,797	53,392	6,481	5,217
3.5 Coefficient of variation (CV)	31.13	32.24	124.96	32.00	30.48
4. Adjusting for unknown residential					
4.1 Sum of weights by residential status before adjustment					
a. Residential – respondent	308,184	305,000	305,083	239,980	189,343
b. Residential – nonrespondent	162,125	171,765	191,808	123,890	103,643
c. Unknown residential status-NA,NM	88,402	94,769	169,432	99313	72,840
d. Nonresidential	633,316	591,620	812,370	593,737	520,720
4.2 Sum of weights before adjustment					
a. Residential – respondent	308,184	305,000	305,083	239,980	189,343
b. Residential – nonrespondent	162,125	171,765	191,808	123,890	103,643
c. Estimated Residential NA, NM	35,071	39,413	58,284	33,645	29,056
4.3 Sum of weights after adjustment					
a. Residential – respondent	308,184	305,000	305,083	239,980	189,343
b. Residential – nonrespondent	197,003	211,028	250,212	157,697	132,520
c. Unknown residential status-NA,NM	0	0	0	0	0
4.4 Sum of weights	505,187	516,028	555,295	397,677	321,863
4.5 Sample size	2,942	2,811	12,507	2,277	1,781
4.6 Coefficient of variation (CV)	52.50	54.52	170.57	64.18	60.58

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	San Bernardino	Riverside	Alameda	Sacramento	Contra Costa
5. Unknown minority screener adjustment					
5.1 Sum of weights before adjustment					
a. Screener minority or other strata	502,832	514,168	544,989	393,790	319,340
b. Non-screener minority	340	398	5,370	584	719
. c. Unknown screener minority	2,015	1,463	4,936	3,303	1,804
5.2 Sum of weights after adjustment					
a. Screener minority or other strata	503,954	515,156	545,945	394,793	319,794
b. Non-screener minority	1,190	1,400	9,453	2,051	2,528
c. Unknown screener minority	0	0	0	0	0
6. Screener nonresponse adjustment					
6.1 Sum of weights before adjustment	505,187	516,028	555,295	397,677	321,863
a. Respondents	308,966	305,451	300,207	240,399	189,077
b. Nonrespondents	196,178	211,104	255,191	156,445	133,245
6.2 Sum of weights after adjustment	505,144	516,555	555,398	396,844	322,322
7. Multiple telephone adjustment					
7.1 Sum of weights before adjustment	505,187	516,028	555,295	397,677	321,863
7.2 Sum of weights after adjustment	484,462	494,685	525,929	383,798	303,982
7.3 Adjustment factor	0.96	0.96	0.95	0.97	0.94
8. Household poststratification					
8.1 Number of completed screeners	2,005	1,901	7,440	1,576	1,199
8.2 Sum of weights before adjustment	484,462	494,685	525,929	383,798	303,982
8.3 Sum of weights after adjustment	528,594	506,218	523,366	453,602	344,129
8.4 Adjustment factor	1.09	1.02	1.00	1.18	1.13
8.5 Coefficient of variation (CV)	24.98	25.88	123.72	29.76	27.81
9 Section G nonresponse adjustment					
9.1 Sum of weights before adjustment	528,594	506,218	523,366	453,602	344,129
a. Completed section-G	383,570	328,295	317,959	338,716	304,568
b. Did not complete section-G	182,293	200,299	188,259	184,650	149,034
9.2 Sum of weights after adjustment	528,594	506,218	523,366	453,602	344,129
9.3 Number of adults completed through section G	2,005	1,901	7,440	1,576	1,199
9.4 Coefficient of variation (CV)	83.92	84.02	168.01	79.76	75.82

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Fresno	San Francisco	Ventura	San Mateo	Kern
1. Base weight					
1.1 Sample size	5,834	12,906	5,722	6,592	4,117
1.2 Sum of weight	650,600	1,068,400	596,500	790,200	507,900
1.3 Coefficient of variation (CV)	6.22	10.25	5.22	8.88	5.56
2. Adjusting for mailable subsampling					
2.1 Sample Size before mailable subsampling					
a. Mail	2,263	4,760	2,313	2,579	1,700
b. Non-mail	1,093	3,589	1,235	1,136	732
c. Nonresidential	2,478	4,557	2,185	2,877	1,685
2.2 Sample Size after mailable subsampling					
a. Mail	2,044	4,379	2,081	2,319	1,539
b. Non-mail	564	2,239	621	581	360
c. Nonresidential	1,882	3,418	1,636	2,157	1,268
2.3 Sample size	4,490	10,036	4,338	5,057	3,167
2.4 Sum of weights after adjustment					
a. Mail	250,221	377,609	241,910	301,004	208,687
b. Non-mail	92,348	273,208	97,430	103,401	67,878
c. Nonresidential	308,032	417,583	257,160	385,796	231,336
2.5 Sum of weights	650,600	1,068,400	596,500	790,200	507,900
2.6 Coefficient of variation (CV)	36.59	35.01	38.90	37.32	36.57
2.7 Observed subsampling rates					
a. Mail subsampling rate	0.903	0.920	0.900	0.899	0.905
b. Non-mail subsampling rate	0.516	0.624	0.503	0.511	0.492
c. Nonresidential subsampling rate	0.759	0.750	0.749	0.750	0.753

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Fresno	San Francisco	Ventura	San Mateo	Kern
3. Refusal Subsampling Adjustment					
3.1 Sum of weights before adjustment					
a. Never refused	541,942	923,000	478,391	666,814	418,183
b. Refused and was selected for refusal conversion	77,393	99,847	83,594	87,032	64,629
c. Refused and was not selected for refusal conversion	31,265	45,552	34,515	36,354	25,088
3.2 Sum of weights after adjustment					
a. Never refused	541,942	923,000	478,391	666,814	418,183
b. Refused and was selected for refusal conversion	108,658	145,400	118,109	123,386	89,717
c. Refused and was not selected for refusal conversion	0	0	0	0	0
3.3 Sum of weights	650,600	1,068,400	596,500	790,200	507,900
3.4 Sample size	4,242	9,542	4,049	4,786	2,988
3.5 Coefficient of variation (CV)	29.12	30.26	31.32	29.82	29.50
4. Adjusting for unknown residential					
4.1 Sum of weights by residential status before adjustment					
a. Residential – respondent	141,814	149,586	131,067	143,570	121,849
b. Residential – nonrespondent	81,547	138,959	79,920	90,176	59,933
c. Unknown residential status-NA,NM	43,176	135,568	50,358	75,241	33,679
d. Nonresidential	382,466	641,593	333,175	478,902	290,445
4.2 Sum of weights before adjustment					
a. Residential – respondent	141,814	149,586	131,067	143,570	121,849
b. Residential – nonrespondent	81,547	138,959	79,920	90,176	59,933
c. Estimated Residential NA, NM	14,864	49,657	18,758	29,246	11,958
4.3 Sum of weights after adjustment					
a. Residential – respondent	141,814	149,586	131,067	143,570	121,849
b. Residential – nonrespondent	96,643	188,763	98,579	119,302	71885
c. Unknown residential status-NA,NM	0	0	0	0	0
4.4 Sum of weights	238,456	338,349	229,646	262,871	193,734
4.5 Sample size	1,514	2,736	1,462	1,493	1,120
4.6 Coefficient of variation (CV)	52.97	82.04	58.85	68.60	51.82

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Fresno	San Francisco	Ventura	San Mateo	Kern
5. Unknown minority screener adjustment					
5.1 Sum of weights before adjustment					
a. Screener minority or other strata	237,093	327,375	228,546	259,476	193,216
b. Non-screener minority	102	1,865	211	545	120
c. Unknown screener minority	1,261	9,110	889	2,850	398
5.2 Sum of weights after adjustment					
a. Screener minority or other strata	237,380	328,822	229,036	260,025	193,431
b. Non-screener minority	356	6,568	737	1,922	421
c. Unknown screener minority	0	0	0	0	0
6. Screener nonresponse adjustment					
6.1 Sum of weights before adjustment	238,456	338,349	229,646	262,871	193,734
a. Respondents	141,998	149,169	131,346	143,573	121,943
b. Nonrespondents	95,737	186,221	98,427	118,373	71,908
6.2 Sum of weights after adjustment	237,736	335,390	229,773	261,947	193,851
7. Multiple telephone adjustment					
7.1 Sum of weights before adjustment	238,456	338,349	229,646	262,871	193,734
7.2 Sum of weights after adjustment	229,738	317,068	217,527	243,990	187,400
7.3 Adjustment factor	0.96	0.94	0.95	0.93	0.97
8. Household poststratification					
8.1 Number of completed screeners	1,011	1,453	962	945	790
8.2 Sum of weights before adjustment	229,738	317,068	217,527	243,990	187,400
8.3 Sum of weights after adjustment	252,940	329,700	243,234	254,103	208,652
8.4 Adjustment factor	1.10	1.04	1.12	1.04	1.11
8.5 Coefficient of variation (CV)	25.77	34.23	29.08	29.09	27.29
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment	252,940	329,700	243,234	254,103	208,652
a. Completed section-G	237,721	160,528	210,957	153,387	161,586
b. Did not complete section-G	106,408	92,412	118,743	89,847	92,517
9.2 Sum of weights after adjustment	252,940	329,700	243,234	254,103	208,652
9.3 Number of adults completed through section G	1,011	1,453	962	945	790
9.4 Coefficient of variation (CV)	81.86	86.37	83.36	81.24	74.05

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
1. Base weight					
1.1 Sample size	4,137	3,916	4,012	4,214	3,512
1.2 Sum of weight	391,800	434,300	310,000	379,700	283,100
1.3 Coefficient of variation (CV)	6.36	4.03	3.56	4.04	4.28
2. Adjusting for mailable subsampling					
2.1 Sample Size before mailable subsampling					
a. Mail	1,904	1,855	1,865	1,597	1,785
b. Non-mail	823	722	754	888	654
c. Nonresidential	1,410	1,339	1,393	1,729	1,073
2.2 Sample Size after mailable subsampling					
a. Mail	1,731	1,727	1,697	1,462	1,626
b. Non-mail	469	449	413	576	438
c. Nonresidential	1,049	1,055	1,048	1,297	805
2.3 Sample size	3,249	3,231	3,158	3,335	2,869
2.4 Sum of weights after adjustment					
a. Mail	178,362	205,722	144,102	139,380	139,808
b. Non-mail	65,608	68,123	46,853	74,069	50,568
c. Nonresidential	147,830	160,455	119,046	166,251	92,724
2.5 Sum of weights	391,800	434,300	310,000	379,700	283,100
2.6 Coefficient of variation (CV)	35.98	30.71	35.63	30.07	31.40
2.7 Observed subsampling rates					
a. Mail subsampling rate	0.909	0.931	0.910	0.915	0.911
b. Non-mail subsampling rate	0.570	0.622	0.548	0.649	0.670
c. Nonresidential subsampling rate	0.744	0.788	0.752	0.750	0.750

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
3. Refusal Subsampling Adjustment					
3.1 Sum of weights before adjustment					
a. Never refused	311,582	352,894	242,957	318,768	221,030
b. Refused and was selected for refusal conversion	55,981	55,048	48,814	39,339	41,628
c. Refused and was not selected for refusal conversion	24,237	26,358	18,230	21,592	20,442
3.2 Sum of weights after adjustment					
a. Never refused	311,582	352,894	242,957	318,768	221,030
b. Refused and was selected for refusal conversion	80,218	81,406	67,043	60,932	62,070
c. Refused and was not selected for refusal conversion	0	0	0	0	0
3.3 Sum of weights	391,800	434,300	310,000	379,700	283,100
3.4 Sample size	3,023	3,014	2,946	3,120	2,637
3.5 Coefficient of variation (CV)	32.83	31.43	30.95	31,73	35.02
4. Adjusting for unknown residential					
4.1 Sum of weights by residential status before adjustment					
a. Residential – respondent	98,974	100,370	83,516	84,607	80,160
b. Residential – nonrespondent	59,322	56,859	43,906	41,784	38,609
c. Unknown residential status-NA,NM	33,625	45,778	23,465	32,222	27,834
d. Nonresidential	198,333	230,317	158,085	219,736	135,553
4.2 Sum of weights before adjustment					
a. Residential – respondent	98,974	100,370	83,516	84,607	80,160
b. Residential – nonrespondent	59,322	56,859	43,906	41,784	38,609
c. Estimated Residential NA, NM	11,775	19,999	9,268	12,201	10,808
4.3 Sum of weights after adjustment					
a. Residential – respondent	98,974	100,370	83,516	84,607	80,160
b. Residential – nonrespondent	71,314	76,837	53,302	54,188	49,289
c. Unknown residential status-NA,NM	0	0	0	0	0
4.4 Sum of weights	170,288	177,207	136,818	138,795	129,449
4.5 Sample size	1,261	1,080	1,251	1,056	1,110
4.6 Coefficient of variation (CV)	55.69	67.54	50.98	64.69	60.46

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano
5. Unknown minority screener adjustment					
5.1 Sum of weights before adjustment					
a. Screener minority or other strata	168,682	176,615	136,231	137,998	129,149
b. Non-screener minority	112	56	141	42	151
c. Unknown screener minority	1,493	535	447	754	150
5.2 Sum of weights after adjustment					
a. Screener minority or other strata	169,033	176,615	136,352	138,106	129,263
b. Non-screener minority	391	199	493	150	529
c. Unknown screener minority	0	0	0	0	0
6. Screener nonresponse adjustment					
6.1 Sum of weights before adjustment	170,288	177,207	136,818	138,795	129,449
a. Respondents	99,108	100,313	83,496	84,672	80,123
b. Nonrespondents	70,317	76,501	53,348	53,584	49,669
6.2 Sum of weights after adjustment	169,425	176,815	136,845	138,255	129,792
7. Multiple telephone adjustment					
7.1 Sum of weights before adjustment	170,288	177,207	136,818	138,795	1129,449
7.2 Sum of weights after adjustment	162,236	168,188	132,643	131,009	125,056
7.3 Adjustment factor	0.95	0.95	0.97	0.94	0.97
8. Household poststratification					
8.1 Number of completed screeners	828	739	858	754	790
8.2 Sum of weights before adjustment	162,236	168,188	132,643	131,009	125,056
8.3 Sum of weights after adjustment	181,629	172,403	145,146	136,622	130,403
8.4 Adjustment factor	1.12	1.03	1.09	1.04	1.04
8.5 Coefficient of variation (CV)	26.91	27.91	22.91	31.61	27.36
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment	181,629	172,403	145,146	136,622	130,403
a. Completed section-G	141,950	115,468	120,545	93,320	93,770
b. Did not complete section-G	66,702	66,161	51,858	51,826	42,852
9.2 Sum of weights after adjustment	181,629	172,403	145,146	136,622	130,403
9.3 Number of adults completed through section G	828	739	858	754	790
9.4 Coefficient of variation (CV)	81.36	74.47	79.26	76.78	78.88

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
1. Base weight					
1.1 Sample size	4,805	4,006	5,215	3,410	3,914
1.2 Sum of weight	282,700	257,600	325,300	233,200	262,700
1.3 Coefficient of variation (CV)	3.10	1.98	2.78	3.19	4.65
2. Adjusting for mailable subsampling					
2.1 Sample Size before mailable subsampling					
a. Mail	1,698	1,621	1,971	1,553	1,561
b. Non-mail	725	793	1,071	746	1,056
c. Nonresidential	2,382	1,592	2,173	1,111	1,297
2.2 Sample Size after mailable subsampling					
a. Mail	1,555	1,621	1,826	1,553	1,424
b. Non-mail	402	594	606	559	695
c. Nonresidential	1,795	1,193	1,619	833	980
2.3 Sample size	3,752	3,408	4,051	2,945	3,099
2.4 Sum of weights after adjustment					
a. Mail	97,965	104,110	123,112	106,052	102,419
b. Non-mail	33,956	51,044	54,921	51,029	66,221
c. Nonresidential	150,779	102,445	147,267	76,120	94,060
2.5 Sum of weights	282,700	257,600	325,300	233,200	262,700
2.6 Coefficient of variation (CV)	31.15	14.44	32.77	14.83	31.88
2.7 Observed subsampling rates					
a. Mail subsampling rate	0.916	1.000	0.926	1.000	0.912
b. Non-mail subsampling rate	0.554	0.749	0.566	0.749	0.658
c. Nonresidential subsampling rate	0.754	0.749	0.745	0.750	0.756

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
3. Refusal Subsampling Adjustment					
3.1 Sum of weights before adjustment					
a. Never refused	242,290	216,317	276,835	188,201	209,750
b. Refused and was selected for refusal conversion	28,003	25,964	33,048	28,208	35,191
c. Refused and was not selected for refusal conversion	12,408	15,318	15,417	16,791	17,759
3.2 Sum of weights after adjustment					
a. Never refused	242,290	216,317	276,835	188,201	209,750
b. Refused and was selected for refusal conversion	40,410	41,283	48,465	44,999	52,950
c. Refused and was not selected for refusal conversion	0	0	0	0	0
3.3 Sum of weights	282,700	257,600	325,300	233,200	262,700
3.4 Sample size	3,559	3,180	3,826	2,710	2,870
3.5 Coefficient of variation (CV)	27.39	32.55	28.84	35.64	34.02
4. Adjusting for unknown residential					
4.1 Sum of weights by residential status before adjustment					
a. Residential – respondent	63,881	57,518	57,633	61,246	65,703
b. Residential – nonrespondent	26,819	30,872	35,739	24,453	31,384
c. Unknown residential status-NA,NM	14,705	26,934	29,763	23,516	30,506
d. Nonresidential	176,936	141,310	201,294	122,585	133,933
4.2 Sum of weights before adjustment					
a. Residential – respondent	63,881	57,518	57,633	61,246	65,703
b. Residential – nonrespondent	26,819	30,872	35,739	24,453	31,384
c. Estimated Residential NA, NM	5,647	11,348	12,318	9,485	10,780
4.3 Sum of weights after adjustment					
a. Residential – respondent	63,881	57,518	57,633	61,246	65,703
b. Residential – nonrespondent	32,499	42,198	48,069	33,842	42,186
c. Unknown residential status-NA,NM	0	0	0	0	0
4.4 Sum of weights	96,380	99,716	105,702	95,088	107,889
4.5 Sample size	1,192	1,100	1,129	989	1,052
4.6 Coefficient of variation (CV)	49.59	71.64	68.77	69.25	70.56

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Tulare	Santa Cruz	Marin	San Luis Obispo	Placer
5. Unknown minority screener adjustment					
5.1 Sum of weights before adjustment					
a. Screener minority or other strata	96,260	99,568	105,434	94,809	107,607
b. Non-screener minority	0	38	53	169	58
c. Unknown screener minority	120	111	215	110	224
5.2 Sum of weights after adjustment					
a. Screener minority or other strata	96,260	99,568	105,567	94,877	107,673
b. Non-screener minority	0	131	183	592	205
c. Unknown screener minority	0	0	0	0	0
6. Screener nonresponse adjustment					
6.1 Sum of weights before adjustment	96,380	99,716	105,702	95,088	107,889
a. Respondents	63,817	57,416	57,646	61,145	65,711
b. Nonrespondents	32,443	42,282	48,104	34,324	42,167
6.2 Sum of weights after adjustment	96,260	99,698	105,750	95,469	107,878
7. Multiple telephone adjustment					
7.1 Sum of weights before adjustment	96,380	99,716	105,702	95,088	107,889
7.2 Sum of weights after adjustment	93,025	93,645	96,575	91,462	103,515
7.3 Adjustment factor	0.97	0.94	0.91	0.96	0.96
8. Household poststratification					
8.1 Number of completed screeners	882	772	749	751	761
8.2 Sum of weights before adjustment	93,025	93,645	96,575	91,462	103,515
8.3 Sum of weights after adjustment	110,385	91,139	100,650	92,739	93,382
8.4 Adjustment factor	1.19	0.97	1.04	1.01	0.90
8.5 Coefficient of variation (CV)	22.34	31.36	34.02	30.91	31.16
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment	110,385	91,139	100,650	92,739	93,382
a. Completed section-G	84,925	73,900	60,415	69,132	62,316
b. Did not complete section-G	45,478	36,485	30,724	31,518	30,423
9.2 Sum of weights after adjustment	110385	91,139	100,650	92,739	93,382
9.3 Number of adults completed through section G	882	772	749	751	761
9.4 Coefficient of variation (CV)	76.41	82.32	77.59	78.51	81.91

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Merced	Butte	Shasta	Yolo	El Dorado
1. Base weight					
1.1 Sample size	3,905	3,208	3,405	3,326	4,004
1.2 Sum of weight	119,100	159,100	139,300	138,500	144,200
1.3 Coefficient of variation (CV)	2.08	41.04	2.76	6.79	2.01
2. Adjusting for mailable subsampling					
2.1 Sample Size before mailable subsampling					
a. Mail	1,773	1,632	1,492	1,530	1,685
b. Non-mail	735	578	763	678	911
c. Nonresidential	1,397	998	1,150	1,118	1,408
2.2 Sample Size after mailable subsampling					
a. Mail	1,655	1,506	1,359	1,416	1,552
b. Non-mail	420	322	406	482	590
c. Nonresidential	1,049	752	865	843	1,052
2.3 Sample size	3,124	2,580	2,630	2,741	3,194
2.4 Sum of weights after adjustment					
a. Mail	54,511	81,479	62,007	61,140	59,754
b. Non-mail	18,464	23,361	24,798	28,140	30,220
c. Nonresidential	46,125	54,260	52,495	49,219	54,226
2.5 Sum of weights	119,100	159,100	139,300	138,500	144,200
2.6 Coefficient of variation (CV)	32.10	34.11	36.97	28.30	30.62
2.7 Observed subsampling rates					
a. Mail subsampling rate	0.933	0.923	0.911	0.925	0.921
b. Non-mail subsampling rate	0.571	0.557	0.532	0.711	0.648
c. Nonresidential subsampling rate	0.751	0.754	0.752	0.754	0.747

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Merced	Butte	Shasta	Yolo	El Dorado
3. Refusal Subsampling Adjustment					
3.1 Sum of weights before adjustment					
a. Never refused	97,522	124,975	110,774	113,417	114,332
b. Refused and was selected for refusal conversion	14,901	23,486	19,789	16,537	19,633
c. Refused and was not selected for refusal conversion	6,677	10,639	8,737	8,545	10,235
3.2 Sum of weights after adjustment					
a. Never refused	97,522	124,975	110,774	113,417	114,332
b. Refused and was selected for refusal conversion	21,578	34,125	28,526	25,083	29,868
c. Refused and was not selected for refusal conversion	0	0	0	0	0
3.3 Sum of weights	119,100	159,100	139,300	138,500	144,200
3.4 Sample size	2,925	2,389	2,443	2,557	2,941
3.5 Coefficient of variation (CV)	30.82	33.37	32.59	34.28	34.49
4. Adjusting for unknown residential					
4.1 Sum of weights by residential status before adjustment					
a. Residential – respondent	31,796	48,573	37,898	36,694	34,852
b. Residential – nonrespondent	16,335	21,251	17,354	16,122	17,904
c. Unknown residential status-NA,NM	8,851	14,266	12,337	12,849	15,559
d. Nonresidential	61,882	74,398	71,151	71,967	75,140
4.2 Sum of weights before adjustment					
a. Residential – respondent	31,796	48,573	37,898	36,694	34,852
b. Residential – nonrespondent	16,335	21,251	17,354	16,122	17,904
c. Estimated Residential NA, NM	3,616	6,310	4,726	4,062	5,923
4.3 Sum of weights after adjustment					
a. Residential – respondent	31,796	48,573	37,898	36,694	34,852
b. Residential – nonrespondent	19,964	27,532	22,098	20,221	23,791
c. Unknown residential status-NA,NM	0	0	0	0	0
4.4 Sum of weights	51,760	76,105	59,996	56,915	58,643
4.5 Sample size	1,213	1,067	981	1,005	1,063
4.6 Coefficient of variation (CV)	52.55	57.16	59.04	64.08	68.25

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Merced	Butte	Shasta	Yolo	El Dorado
5. Unknown minority screener adjustment					
5.1 Sum of weights before adjustment					
a. Screener minority or other strata	51,683	75,951	59,823	56,407	58,619
b. Non-screener minority	21	21	30	136	23
c. Unknown screener minority	55	133	143	372	0
5.2 Sum of weights after adjustment					
a. Screener minority or other strata	51,757	75,951	59,823	56,522	58,619
b. Non-screener minority	74	72	106	478	81
c. Unknown screener minority	0	0	0	0	0
6. Screener nonresponse adjustment					
6.1 Sum of weights before adjustment	51,760	76,105	59,996	56,915	58,643
a. Respondents	31,849	48,498	37,868	36,585	34,828
b. Nonrespondents	19,983	27,525	22,060	20,416	23,872
6.2 Sum of weights after adjustment	51,831	76,023	59,929	57,001	58,700
7. Multiple telephone adjustment					
7.1 Sum of weights before adjustment	51,760	76,105	59,996	56,915	58,643
7.2 Sum of weights after adjustment	50,344	72,882	57,496	54,666	55,871
7.3 Adjustment factor	0.97	0.96	0.96	0.96	0.95
8. Household poststratification					
8.1 Number of completed screeners	847	789	718	733	751
8.2 Sum of weights before adjustment	50,344	72,882	57,496	54,666	55,871
8.3 Sum of weights after adjustment	63,815	79,566	63,426	59,375	58,939
8.4 Adjustment factor	1.27	1.09	1.10	1.09	1.05
8.5 Coefficient of variation (CV)	24.08	26.71	24.32	30.78	29.89
9. Section G nonresponse adjustment					
9.1 Sum of weights before adjustment	63,815	79,566	63,426	59,375	58,939
a. Completed section-G	61,716	40,881	57,577	44,793	41,459
b. Did not complete section-G	31,666	22,934	21,989	18,633	17,916
9.2 Sum of weights after adjustment	63,815	79,566	63,426	59,375	58,939
9.3 Number of adults completed through section G	847	789	718	733	751
9.4 Coefficient of variation (CV)	80.16	69.45	70.61	73.96	78.08

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

		Imperial	Napa	Kings	Madera	Monterey, San Benito
1.	Base weight					
1.1	Sample size	4,102	3,904	4,004	4,202	5,014
1.2	Sum of weight	84,300	114,471	70,400	87,300	388,200
1.3	Coefficient of variation (CV)	1.53	2.29	1.73	1.09	3.68
2.	Adjusting for mailable subsampling					
2.1	Sample Size before mailable subsampling					
a.	Mail	1,778	1,686	1,763	1,545	1,846
b.	Non-mail	705	851	837	1,158	896
c.	Nonresidential	1,619	1,367	1,404	1,499	2,272
2.2	Sample Size after mailable subsampling					
a.	Mail	1,639	1,686	1,622	1,427	1,701
b.	Non-mail	469	640	470	622	495
c.	Nonresidential	1,218	1,025	1,054	1,129	1,703
2.3	Sample size	3,326	3,351	3,146	3,178	3,899
2.4	Sum of weights after adjustment					
a.	Mail	35,481	49,434	31,264	33,164	141,829
b.	Non-mail	13,710	24,949	12,021	19,256	55,384
c.	Nonresidential	35,109	40,088	27,115	34,880	190,987
2.5	Sum of weights	84,300	114,471	70,400	87,300	388,200
2.6	Coefficient of variation (CV)	28.31	14.47	34.51	37.00	32.61
2.7	Observed subsampling rates					
a.	Mail subsampling rate	0.922	1.000	0.920	0.924	0.921
b.	Non-mail subsampling rate	0.665	0.752	0.562	0.537	0.552
c.	Nonresidential subsampling rate	0.752	0.750	0.751	0.753	0.750

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Imperial	Napa	Kings	Madera	Monterey, San Benito
3. Refusal Subsampling Adjustment					
3.1 Sum of weights before adjustment					
a. Never refused	70,619	92,138	56,736	71,456	329,338
b. Refused and was selected for refusal conversion	8,737	13,230	9,424	11,644	41,645
c. Refused and was not selected for refusal conversion	4,944	9,102	4,241	4,200	17,217
3.2 Sum of weights after adjustment					
a. Never refused	70,619	92,138	56,736	71,456	329,338
b. Refused and was selected for refusal conversion	13,681	22,333	13,664	15,844	58,862
c. Refused and was not selected for refusal conversion	0	0	0	0	0
3.3 Sum of weights	84,300	114,471	70,400	87,300	388,200
3.4 Sample size	3,106	3,057	2,932	3,005	3,698
3.5 Coefficient of variation (CV)	32.08	37.89	31.63	28.08	27.65
4. Adjusting for unknown residential					
4.1 Sum of weights by residential status before adjustment					
a. Residential – respondent	21,230	26,156	18,461	22,036	77,204
b. Residential – nonrespondent	10,577	14,879	10,030	11,385	42,922
c. Unknown residential status-NA,NM	7,324	15,241	5,501	5,458	33,506
d. Nonresidential	45,040	57,426	36,237	48,208	233,678
4.2 Sum of weights before adjustment					
a. Residential – respondent	21,230	26,156	18,461	22,036	77,204
b. Residential – nonrespondent	10,577	14,879	10,030	11,385	42,922
c. Estimated Residential NA, NM	2,403	5,298	2,258	1,989	12,650
4.3 Sum of weights after adjustment					
a. Residential – respondent	21,230	26,156	18,461	22,036	77,204
b. Residential – nonrespondent	12,987	20,191	12,278	13,380	55,586
c. Unknown residential status-NA,NM	0	0	0	0	0
4.4 Sum of weights	34,216	46,347	30,738	35,416	132,790
4.5 Sample size	1,198	1,074	1,206	1,187	1,195
4.6 Coefficient of variation (CV)	59.65	79.80	54.23	48.16	62.53

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

		Imperial	Napa	Kings	Madera	Monterey, San Benito
5.	Unknown minority screener adjustment					
	5.1 Sum of weights before adjustment					
	a. Screener minority or other strata	34,189	46,205	30,695	35,392	132,383
	b. Non-screener minority	0	0	0	0	68
	c. Unknown screener minority	28	143	43	24	339
	5.2 Sum of weights after adjustment					
	a. Screener minority or other strata	34,189	46,205	30,695	35,392	132,549
	b. Non-screener minority	0	0	0	0	240
	c. Unknown screener minority	0	0	0	0	0
6.	Screener nonresponse adjustment					
	6.1 Sum of weights before adjustment	34,216	46,347	30,738	35,416	132,790
	a. Respondents	21,230	26,156	18,429	22,036	77,302
	b. Nonrespondents	12,959	20,048	12,266	13,356	55,487
	6.2 Sum of weights after adjustment	34,189	46,205	30,695	35,392	132,789
7.	Multiple telephone adjustment					
	7.1 Sum of weights before adjustment	34,216	46,347	30,738	35,416	132,790
	7.2 Sum of weights after adjustment	33,220	43,243	29,875	34,283	126,020
	7.3 Adjustment factor	0.97	0.93	0.97	0.97	0.95
8.	Household poststratification					
	8.1 Number of completed screeners	857	756	837	828	810
	8.2 Sum of weights before adjustment	33,220	43,243	29,875	34,283	126,020
	8.3 Sum of weights after adjustment	39,384	45,402	34,418	36,155	137,121
	8.4 Adjustment factor	1.19	1.05	1.15	1.05	1.09
	8.5 Coefficient of variation (CV)	26.37	37.69	24.10	22.61	26.62
9.	Section G nonresponse adjustment					
	9.1 Sum of weights before adjustment	39,384	45,402	34,418	36,155	137,121
	a. Completed section-G	39,137	25,111	30,699	22,254	22,528
	b. Did not complete section-G	19,802	14,273	14,703	12,164	13,627
	9.2 Sum of weights after adjustment	39,384	45,402	34,418	36,155	137,121
	9.3 Number of adults completed through section G	857	756	837	828	810
	9.4 Coefficient of variation (CV)	81.64	86.47	78.79	81.40	79.31

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Del Norte, Homboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	Colusa, Glen, Tehama
1. Base weight				
1.1 Sample size	4,103	3,503	3,204	2,900
1.2 Sum of weight	148,400	95,100	129,967	83,500
1.3 Coefficient of variation (CV)	1.86	1.07	1.37	1.36
2. Adjusting for mailable subsampling				
2.1 Sample Size before mailable subsampling				
a. Mail	1,374	1,163	1,290	1,226
b. Non-mail	804	651	607	621
c. Nonresidential	1,925	1,689	1,307	1,053
2.2 Sample Size after mailable subsampling				
a. Mail	1,242	1,026	1,146	1,077
b. Non-mail	444	339	289	302
c. residential	1,443	1,268	980	786
2.3 Sample size	3,129	2,633	2,415	2,165
2.4 Sum of weights after adjustment				
a. Mail	49,030	30,812	52,470	35,552
b. Non-mail	23,292	13,534	17,653	13,178
c. Nonresidential	76,078	50,754	59,844	34,771
2.5 Sum of weights	148,400	95,100	129,967	83,500
2.6 Coefficient of variation (CV)	33.80	36.00	38.90	41.27
2.7 Observed subsampling rates				
a. Mail subsampling rate	0.904	0.882	0.888	0.878
b. Non-mail subsampling rate	0.552	0.521	0.476	0.486
c. Nonresidential subsampling rate	0.750	0.751	0.750	0.746

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Del Norte, Homboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	Colusa, Glen, Tehama
3. Refusal Subsampling Adjustment				
3.1 Sum of weights before adjustment				
a. Never refused	128,134	81,676	107,418	67,658
b. Refused and was selected for refusal conversion	14,499	9,808	15,696	11,286
c. Refused and was not selected for refusal conversion	5,767	3,615	6,853	4,556
3.2 Sum of weights after adjustment				
a. Never refused	128,134	81,676	107,418	67,658
b. Refused and was selected for refusal conversion	20,266	13,424	22,549	15,842
c. Refused and was not selected for refusal conversion	0	0	0	0
3.3 Sum of weights	148,400	985,100	129,967	83,500
3.4 Sample size	2,986	2,514	2,266	2,028
3.5 Coefficient of variation (CV)		25.48	30.04	30.23
4. Adjusting for unknown residential				
4.1 Sum of weights by residential status before adjustment				
a. Residential – respondent	32,165	19,554	30,471	24,617
b. Residential – nonrespondent	14,034	8,002	13,995	9,431
c. Unknown residential status-NA,NM	8,810	6,048	10,984	5,717
d. Nonresidential	93,173	61,334	74,148	43,628
4.2 Sum of weights before adjustment				
a. Residential – respondent	32,165	19,554	30,471	24,617
b. Residential – nonrespondent	14,034	8,002	13,995	9,431
c. Estimated Residential NA, NM	3,790	2,322	4,882	2,130
4.3 Sum of weights after adjustment				
a. Residential – respondent	32,165	19,554	30,471	24,617
b. Residential – nonrespondent	17,861	10,324	18,816	11,561
c. Unknown residential status-NA,NM	0	0	0	0
4.4 Sum of weights	50,026	29,878	49,288	36,178
4.5 Sample size	983	761	798	853
4.6 Coefficient of variation (CV)	53.78	56.37	62.45	50.42

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Del Norte, Homboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	Colusa, Glen, Tehama
5. Unknown minority screener adjustment				
5.1 Sum of weights before adjustment				
a. Screener minority or other strata	49,897	29,859	49,097	36,178
b. Non-screener minority	38	19	62	0
c. Unknown screener minority	91	0	129	0
5.2 Sum of weights after adjustment				
a. Screener minority or other strata	49,897	29,859	49,097	36,178
b. Non-screener minority	133	68	218	0
c. Unknown screener minority	0	0	0	0
6. Screener nonresponse adjustment				
6.1 Sum of weights before adjustment	50,026	29,878	49,288	36,178
a. Respondents	32,128	19,535	30,410	24,617
b. Nonrespondents	17,903	10,392	18,906	11,561
6.2 Sum of weights after adjustment	50,030	29,927	49,316	36,178
7. Multiple telephone adjustment				
7.1 Sum of weights before adjustment	50,026	29,878	49,288	36,178
7.2 Sum of weights after adjustment	48,328	29,050	47,616	35,108
7.3 Adjustment factor	0.97	0.97	0.97	0.97
8. Household poststratification				
8.1 Number of completed screeners	728	568	579	650
8.2 Sum of weights before adjustment	48,328	29,050	47,616	35,108
8.3 Sum of weights after adjustment	60,408	37,552	57,240	36,282
8.4 Adjustment factor	1.25	1.29	1.20	1.03
8.5 Coefficient of variation (CV)	20.69	22.76	25.36	22.37
9. Section G nonresponse adjustment				
9.1 Sum of weights before adjustment	60,408	37,552	57,240	36,282
a. Completed section-G	90,242	44,107	27,603	40,344
b. Did not complete section-G	46,879	16,301	9,949	16,896
9.2 Sum of weights after adjustment	60,408	37,552	57,240	36,282
9.3 Number of adults completed through section G	728	568	579	650
9.4 Coefficient of variation (CV)	65.86	66.33	69.43	75.77

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1. Base weight			
1.1 Sample size	3,103	3,103	3,701
1.2 Sum of weight	100,700	135,900	226,000
1.3 Coefficient of variation (CV)	1.93	2.14	0.95
2. Adjusting for mailable subsampling			
2.1 Sample Size before mailable subsampling			
a a. Mail	1,333	1,182	1,326
b b. Non-mail	661	700	897
c c. Nonresidential	1,109	1,221	1,478
2.2 Sample Size after mailable subsampling			
a. Mail	1,188	1,182	1,191
b. Non-mail	315	527	460
c. Nonresidential	826	911	1,105
2.3 Sample size	2,329	2,620	2,756
2.4 Sum of weights after adjustment			
a. Mail	44,033	51,700	82,014
b. Non-mail	15,565	30,654	42,119
c. Nonresidential	41,102	53,546	101,867
2.5 Sum of weights	100,700	135,900	226,000
2.6 Coefficient of variation (CV)	40.57	14.38	38.63
2.7 Observed subsampling rates			
a. Mail subsampling rate	0.891	1.000	0.898
b. Non-mail subsampling rate	0.477	0.753	0.513
c. Nonresidential subsampling rate	0.745	0.746	0.748

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
3. Refusal Subsampling Adjustment			
3.1 Sum of weights before adjustment			
a. Never refused	81,948	112,482	187,890
b. Refused and was selected for refusal conversion	13,771	14,330	26,881
c. Refused and was not selected for refusal conversion	4,981	9,088	11,230
3.2 Sum of weights after adjustment			
a. Never refused	81,948	112,482	187,890
b. Refused and was selected for refusal conversion	18,752	23,418	38,110
c. Refused and was not selected for refusal conversion	0	0	0
3.3 Sum of weights	100,700	15,900	226,000
3.4 Sample size	2,196	2,423	2,598
3.5 Coefficient of variation (CV)	28.83	34.55	28.82
4. Adjusting for unknown residential			
4.1 Sum of weights by residential status before adjustment			
a. Residential – respondent	29,250	30,460	45,594
b. Residential – nonrespondent	11,601	14,400	24,407
c. Unknown residential status-NA,NM	5,708	16,646	22,868
d. Nonresidential	53,775	73,436	132,477
4.2 Sum of weights before adjustment			
a. Residential – respondent	29,250	30,460	45,594
b. Residential – nonrespondent	11,601	14,400	24,407
c. Estimated Residential NA, NM	2,564	6,923	9,684
4.3 Sum of weights after adjustment			
a. Residential – respondent	29,2580	30,460	45,594
b. Residential – nonrespondent	14,189	21,352	34,091
c. Unknown residential status-NA,NM	0	0	0
4.4 Sum of weights	43,439	51,812	79,686
4.5 Sample size	931	790	818
4.6 Coefficient of variation (CV)	45.88	80.85	71.00

Table B–1. Household weighting for the combined RDD and surname list samples by stratum (continued)

	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
5. Unknown minority screener adjustment			
5.1 Sum of weights before adjustment			
a. Screener minority or other strata	43,331	51,698	79,686
b. Non-screener minority	0	27	0
c. Unknown screener minority	108	87	0
5.2 Sum of weights after adjustment			
a. Screener minority or other strata	43,331	51,698	79,686
b. Non-screener minority	0	96	0
c. Unknown screener minority	0	0	0
6. Screener nonresponse adjustment			
6.1 Sum of weights before adjustment	43,439	51,812	79,686
a. Respondents	29,250	30,433	45,594
b. Nonrespondents	14,081	21,361	34,091
6.2 Sum of weights after adjustment	43,331	51,794	79,686
7. Multiple telephone adjustment			
7.1 Sum of weights before adjustment	43,439	51,812	79,686
7.2 Sum of weights after adjustment	41,561	48,991	77,158
7.3 Adjustment factor	0.96	0.95	0.97
8. Household poststratification			
8.1 Number of completed screeners	697	581	575
8.2 Sum of weights before adjustment	41,561	48,991	77,158
8.3 Sum of weights after adjustment	47,568	47,414	70,168
8.4 Adjustment factor	1.14	0.97	0.91
8.5 Coefficient of variation (CV)	23.92	31.94	29.25
9. Section G nonresponse adjustment			
9.1 Sum of weights before adjustment	47,568	47,414	70,168
a. Completed section-G	24,092	32,004	33,295
b. Did not complete section-G	12,190	15,564	14,119
9.2 Sum of weights after adjustment	47,568	47,414	70,168
9.3 Number of adults completed through section G	697	581	575
9.4 Coefficient of variation (CV)	75.00	77.62	72.33

Table B-2. 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	66,657	17,770	3,701	3,674	2,042
1.2 Sum of weights	23,935,170	6,723,805	2,006,506	2,020,508	1,209,569
1.3 Coefficient of variation (CV)	82.37	63.29	64.97	81.31	57.57
2. Nonresponse adjustment					
2.1 Number of completed adult interviews	42,044	10,350	2,310	2,231	1,340
2.2 Sum of weights before adjustment	23,935,170	6,723,805	2,006,506	2,020,508	1,209,569
a. Eligible respondents	14,000,796	3,591,364	1,192,554	1,140,829	756,058
b. Ineligibles	586,936	199,006	40,090	51,275	32,443
c. Nonrespondents	9,347,438	2,933,436	773,862	828,404	421,068
2.3 Sum of weights after adjustment	23,935,169	6,723,805	2,006,506	2,020,508	1,209,569
a. Eligible respondents	22,922,619	6,358,062	1,939,066	1,929,235	1,157,012
b. Ineligibles	1,012,550	365,742	67,440	91,273	52,557
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	85.96	65.19	66.63	84.78	57.83
2.5 Mean adjustment factor	1.71	1.87	1.68	1.77	1.60
3. Trimming adjustment					
3.1 Number of trimmed records	41	0	2	3	1
3.2 Sum of weights before trimming adjustment	22,922,619	6,378,895	1,950,206	1,882,639	1,192,723
3.3 Sum of weights after trimming adjustment	22,891,793	6,378,895	1,948,772	1,875,398	1,191,819
4. Raking nonresponse adjustment					
4.1 Number of completed adult interviews	42,044	10,363	2,319	2,186	1,395
4.2 Sum of weights after adjustment	25,597,062	7,104,754	2,171,885	2,164,412	1,255,764
4.3 Mean adjustment factor	1.12	1.11	1.11	1.15	1.05
4.4 Coefficient of variation (CV)	98.31	74.18	82.34	93.83	69.60
4.5 Mean weight	608.82	685.59	936.56	990.12	900.19

Table B-2. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	San Bernardino	Riverside	Hayward	Oakward	Remainder of Amameda
1. Adult initial weights					
1.1 Number of sampled adults	2,005	1,901	2,731	3,006	1,703
1.2 Sum of weights	1,132,869	1,069,063	259,313	314,602	499,549
1.3 Coefficient of variation (CV)	51.54	48.80	207.59	140.90	82.77
2. Nonresponse adjustment					
2.1 Number of completed adult interviews	1,238	1,180	1,629	1,975	1,130
2.2 Sum of weights before adjustment	1,132,869	1,069,063	259,313	314,602	499,549
a. Eligible respondents	656,429	614,495	146,983	195,742	307,540
b. Ineligibles	29,041	21,517	6,545	7,129	13,031
c. Nonrespondents	447,398	433,051	105,786	111,732	178,978
2.3 Sum of weights after adjustment	1,132,869	1,069,063	259,313	314,602	499,549
a. Eligible respondents	1,081,372	1,032,767	247,913	303,232	478,118
b. Ineligibles	51,496	36,296	11,400	11,370	21,431
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	56.21	54.84	215.92	131.81	88.37
2.5 Mean adjustment factor	1.73	1.74	1.76	1.61	1.62
3. Trimming adjustment					
3.1 Number of trimmed records	0	2	4	2	0
3.2 Sum of weights before trimming adjustment	1,082,527	1,034,345	57,796	254,556	687,023
3.3 Sum of weights after trimming adjustment	1,082,527	1,032,928	56,548	251,500	687,023
4. Raking nonresponse adjustment					
4.1 Number of completed adult interviews	1,244	1,186	788	1,853	2,006
4.2 Sum of weights after adjustment	1,255,477	1,213,544	102,369	304,685	700,859
4.3 Mean adjustment factor	1.16	1.17	1.81	1.21	1.02
4.4 Coefficient of variation (CV)	76.83	66.96	124.11	99.35	129.65
4.5 Mean weight	1,009.23	1,023.22	129.91	164.43	349.38

Table B-2. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Sacramento	Contra Costa	Fresno	San Francisco	Ventura
1. Adult initial weights					
1.1 Number of sampled adults	1,576	1,199	1,011	1,453	962
1.2 Sum of weights	900,686	692,603	531,495	629,334	507,445
1.3 Coefficient of variation (CV)	54.73	50.76	52.10	57.81	54.99
2. Nonresponse adjustment					
2.1 Number of completed adult interviews	1,062	820	626	917	617
2.2 Sum of weights before adjustment	900,686	692,603	531,495	629,334	507,445
a. Eligible respondents	558,293	449,459	320,610	369,116	297,553
b. Ineligibles	14,361	15,096	11,073	12,653	13,827
c. Nonrespondents	328,032	228,048	199,812	247,566	196,066
2.3 Sum of weights after adjustment	900,686	692,603	531,495	629,334	507,445
a. Eligible respondents	877,598	668,820	513,303	607,538	483,954
b. Ineligibles	23,088	23,784	18,192	21,796	23,492
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	58.83	53.54	56.16	54.78	58.24
2.5 Mean adjustment factor	1.61	1.54	1.66	1.70	1.71
3. Trimming adjustment					
3.1 Number of trimmed records	0	0	2	1	0
3.2 Sum of weights before trimming adjustment	874,161	693,169	508,081	591,801	493,817
3.3 Sum of weights after trimming adjustment	874,161	693,169	506,309	591,400	493,817
4. Raking nonresponse adjustment					
4.1 Number of completed adult interviews	1,061	897	630	904	630
4.2 Sum of weights after adjustment	953,384	732,312	579,722	650,713	568,986
4.3 Mean adjustment factor	1.09	1.06	1.15	1.10	1.15
4.4 Coefficient of variation (CV)	75.77	71.38	75.55	65.31	88.47
4.5 Mean weight	898.57	816.40	920.19	719.82	903.15

Table B-2. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	San Mateo	Kern	San Joaquin	Sonoma	Stanislaus
1. Adult initial weights					
1.1 Number of sampled adults	945	790	828	739	858
1.2 Sum of weights	519,056	429,150	384,014	331,276	318,133
1.3 Coefficient of variation (CV)	56.61	48.56	52.15	51.29	50.13
2. Nonresponse adjustment					
2.1 Number of completed adult interviews	609	537	521	507	549
2.2 Sum of weights before adjustment	519,056	429,150	384,014	331,276	318,133
a. Eligible respondents	305,977	275,460	223,255	215,162	195,107
b. Ineligibles	20,178	8,416	7,053	9,968	5,396
c, Nonrespondents	192,901	145,274	153,706	106,145	117,631
2.3 Sum of weights after adjustment	519,056	429,150	384,014	331,276	318,133
a. Eligible respondents	485,271	415,418	371,958	316,350	309,242
b. Ineligibles	33,785	13,732	12,055	14,926	8,891
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	61.31	53.05	57.24	52.46	56.82
2.5 Mean adjustment factor	1.70	1.56	1.72	1.54	1.63
3. Trimming adjustment					
3.1 Number of trimmed records	3	2	2	0	0
3.2 Sum of weights before trimming adjustment	474,006	423,001	376,212	318,930	296,797
3.3 Sum of weights after trimming adjustment	470,473	420,856	374,121	318,930	296,797
4. Raking nonresponse adjustment					
4.1 Number of completed adult interviews	596	549	523	519	531
4.2 Sum of weights after adjustment	538,355	475,306	421,546	352,533	335,463
4.3 Mean adjustment factor	1.14	1.13	1.13	1.11	1.13
4.4 Coefficient of variation (CV)	77.30	76.98	80.09	64.50	71.61
4.5 Mean weight	903.28	865.77	806.01	679.25	631.76

Table B-2. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Santa Barbara	Salano	Tulare	Santa Cruz	Marin
1. Adult initial weights					
1.1 Number of sampled adults	754	790	882	772	749
1.2 Sum of weights	277,651	268,067	234,821	187,918	184,857
1.3 Coefficient of variation (CV)	67.18	50.33	47.52	53.47	56.74
2. Nonresponse adjustment					
2.1 Number of completed adult interviews	504	510	575	512	521
2.2 Sum of weights before adjustment	277,651	268,067	234,821	187,918	184,857
a. Eligible respondents	174,479	159,067	147,896	117,557	117,084
b. Ineligibles	7,479	6,463	6,244	4,161	5,198
c. Nonrespondents	95,693	102,537	80,681	66,199	62,575
2.3 Sum of weights after adjustment	277,651	268,067	234,821	187,918	184,857
a. Eligible respondents	265,781	256,803	225,523	181,432	176,585
b. Ineligibles	11,869	11,263	9,298	6,486	8,272
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	67.43	52.37	53.51	53.63	53.25
2.5 Mean adjustment factor	1.59	1.69	1.59	1.60	1.58
3. Trimming adjustment					
3.1 Number of trimmed records	1	0	0	1	0
3.2 Sum of weights before trimming adjustment	260,529	250,988	231,200	165,867	177,299
3.3 Sum of weights after trimming adjustment	259,465	250,988	231,200	165,040	177,299
4. Raking nonresponse adjustment					
4.1 Number of completed adult interviews	497	503	582	480	522
4.2 Sum of weights after adjustment	293,634	286,400	257,862	190,857	187,719
4.3 Mean adjustment factor	1.13	1.14	1.12	1.16	1.06
4.4 Coefficient of variation (CV)	87.56	78.81	68.70	61.73	68.40
4.5 Mean weight	590.81	569.38	443.06	397.62	359.62

Table B-2. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	San Luis Obispo	Placer	Merced	Butte	Shasta
1. Adult initial weights					
1.1 Number of sampled adults	751	761	847	789	718
1.2 Sum of weights	179,955	179,915	134,027	143,977	119,114
1.3 Coefficient of variation (CV)	52.67	51.26	48.92	46.34	47.95
2. Nonresponse adjustment					
2.1 Number of completed adult interviews	503	507	520	564	506
2.2 Sum of weights before adjustment	179,955	179,915	134,027	143,977	119,114
a. Eligible respondents	114,511	112,194	76,340	98,373	77,648
b. Ineligibles	3,389	1,900	1,850	2,452	2,650
c. Nonrespondents	62,056	65,821	55,837	43,152	38,816
2.3 Sum of weights after adjustment	179,955	179,915	134,027	143,977	119,114
a. Eligible respondents	174,607	176,697	130,939	140,078	115,084
b. Ineligibles	5,348	3,218	3,087	3,899	4,030
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	56.13	58.96	51.51	50.98	50.56
2.5 Mean adjustment factor	1.57	1.60	1.76	1.46	1.53
3. Trimming adjustment					
3.1 Number of trimmed records	0	2	2	3	0
3.2 Sum of weights before trimming adjustment	176,963	179,125	143,181	140,858	119,486
3.3 Sum of weights after trimming adjustment	179,963	178,294	142,854	140,144	119,486
4. Raking nonresponse adjustment					
4.1 Number of completed adult interviews	506	513	537	567	537
4.2 Sum of weights after adjustment	187,014	209,564	153,043	158,595	129,872
4.3 Mean adjustment factor	1.06	1.18	1.07	1.13	1.09
4.4 Coefficient of variation (CV)	61.65	64.98	73.26	64.29	67.54
4.5 Mean weight	369.59	408.51	285.00	279.71	241.85

Table B-2. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Yolo	El Dorado	Imperial	Napa	Kings
1. Adult initial weights					
1.1 Number of sampled adults	733	751	857	756	837
1.2 Sum of weights	122,069	116,807	85,872	88,685	73,173
1.3 Coefficient of variation (CV)	63.03	48.09	50.20	57.95	46.87
2. Nonresponse adjustment					
2.1 Number of completed adult interviews	517	503	529	505	531
2.2 Sum of weights before adjustment	122,069	116,807	85,872	88,685	73,173
a. Eligible respondents	80,097	73,599	52,350	56,758	44,428
b. Ineligibles	1,347	2,451	1,361	1,891	1,123
c. Nonrespondents	40,625	40,756	32,161	30,036	27,622
2.3 Sum of weights after adjustment	122,069	116,807	85,872	88,685	73,173
a. Eligible respondents	119,747	112,955	83,541	85,689	71,223
b. Ineligibles	2,322	3,851	2,332	2,996	1,950
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	62.32	50.93	56.05	63.92	53.21
2.5 Mean adjustment factor	1.52	1.59	1.64	1.56	1.65
3. Trimming adjustment					
3.1 Number of trimmed records	1	0	0	1	0
3.2 Sum of weights before trimming adjustment	119,223	115,143	85,176	91,563	70,914
3.3 Sum of weights after trimming adjustment	119,025	115,143	85,176	91,354	70,914
4. Raking nonresponse adjustment					
4.1 Number of completed adult interviews	514	506	528	513	528
4.2 Sum of weights after adjustment	130,231	125,671	101,130	94,461	84,305
4.3 Mean adjustment factor	1.09	1.09	1.19	1.03	1.19
4.4 Coefficient of variation (CV)	72.40	56.02	71.58	84.10	69.64
4.5 Mean weight	253.37	248.36	191.53	184.13	159.67

Table B-2. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino
1. Adult initial weights					
1.1 Number of sampled adults	828	810	728	568	579
1.2 Sum of weights	77,742	301,739	113,110	71,207	107,284
1.3 Coefficient of variation (CV)	48.54	52.47	46.58	43.56	45.83
2. Nonresponse adjustment					
2.1 Number of completed adult interviews	512	520	529	419	409
2.2 Sum of weights before adjustment	77,742	301,739	113,110	71,207	107,284
a. Eligible respondents	45,580	184,636	79,199	50,909	72,355
b. Ineligibles	1,683	9,226	1,593	818	622
c. Nonrespondents	30,479	107,877	32,319	19,480	34,307
2.3 Sum of weights after adjustment	77,742	301,739	113,110	71,207	107,284
a. Eligible respondents	74,852	287,092	110,802	70,098	106,296
b. Ineligibles	2,890	14,647	2,309	1,110	988
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	57.07	56.19	49.24	45.79	49.57
2.5 Mean adjustment factor	1.71	1.63	1.43	1.40	1.48
3. Trimming adjustment					
3.1 Number of trimmed records	0	0	1	0	2
3.2 Sum of weights before trimming adjustment	74,013	297,256	109,312	71,989	102,578
3.3 Sum of weights after trimming adjustment	74,013	297,256	109,188	71,989	102,166
4. Raking nonresponse adjustment					
4.1 Number of completed adult interviews	506	542	525	423	396
4.2 Sum of weights after adjustment	87,582	327,532	113,513	70,269	112,316
4.3 Mean adjustment factor	1.18	1.10	1.04	0.98	1.10
4.4 Coefficient of variation (CV)	73.05	65.52	58.63	57.64	55.47
4.5 Mean weight	173.09	604.30	216.21	166.12	283.63

Table B-2. Extended interview weighting for adult interview by stratum (combined RDD and surname list samples) (continued)

	Colusa, Glen, Tehama	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1. Adult initial weights				
1.1 Number of sampled adults	650	697	581	575
1.2 Sum of weights	69,965	97,094	89,128	132,015
1.3 Coefficient of variation (CV)	45.24	48.82	50.92	48.26
2. Nonresponse adjustment				
2.1 Number of completed adult interviews	425	460	403	412
2.2 Sum of weights before adjustment	69,965	97,094	89,128	132,015
a. Eligible respondents	43,547	61,716	58,815	89,676
b. Ineligibles	877	1,675	144	2,241
c. Nonrespondents	25,541	33,704	30,170	40,098
2.3 Sum of weights after adjustment	69,965	97,094	89,128	132,015
a. Eligible respondents	68,497	94,422	88,910	128,737
b. Ineligibles	1468	2,672	219	3,278
c. Nonrespondents	0	0	0	0
2.4 Coefficient of variation (CV)	51.12	53.85	50.31	52.26
2.5 Mean adjustment factor	1.61	1.57	1.52	1.47
3. Trimming adjustment				
3.1 Number of trimmed records	0	0	1	2
3.2 Sum of weights before trimming adjustment	64,408	93,141	86,409	125,313
3.3 Sum of weights after trimming adjustment	64,408	93,141	86,247	124,598
4. Raking nonresponse adjustment				
4.1 Number of completed adult interviews	397	451	390	401
4.2 Sum of weights after adjustment	76,040	102,355	94,950	140,078
4.3 Mean adjustment factor	1.18	1.10	1.10	1.12
4.4 Coefficient of variation (CV)	61.70	60.12	60.19	65.62
4.5 Mean weight	191.54	226.95	243.46	349.32

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples)

	All strata	Los Angeles	San Diego	Orange	Santa Clara
1. Children initial weights					
1.1 Number of sampled children	10,440	2,651	551	587	355
1.2 Sum of weights	5,917,022	1,721,072	473,173	493,055	273,974
1.3 Coefficient of variation (CV)	103.81	84.86	89.78	96.46	74.33
2. Nonresponse adjustment					
2.1 Number of completed child interviews	8,526	2,112	457	466	279
2.2 Sum of weights before adjustment	5,917,022	1,721,072	473,173	493,055	273,974
a. Eligible respondents	4,777,683	1,370,015	394,772	378,381	217,559
b. Ineligibles	48,073	13,183	4,242	5,000	4,460
c. Nonrespondents	1,091,266	337,873	74,160	109,673	51,955
2.3 Sum of weights after adjustment	5,917,022	1,721,072	473,173	493,055	273,974
a. Eligible respondents	5,857,483	1,704,585	468,379	486,567	268,320
b. Ineligibles	59,539	16,487	4,794	6,487	5,653
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	103.27	86.36	91.18	95.64	75.53
2.5 Mean adjustment factor	1.24	1.26	1.20	1.30	1.26
3. Trimming adjustment					
3.1 Number of trimmed records	53	1	1	1	1
3.2 Sum of weights before trimming adjustment	5,857,483	1,715,596	474,666	464,061	279,557
3.3 Sum of weights after trimming adjustment	5,799,965	1,709,513	471,809	458,709	278,164
4. Raking nonresponse adjustment					
4.1 Number of completed child interviews	8,526	2,115	461	454	290
4.2 Sum of weights after adjustment	6,228,712	1,852,707	461,917	529,311	301,352
4.3 Mean adjustment factor	1.07	1.08	0.98	1.15	1.08
4.4 Coefficient of variation (CV)	102.16	86.57	87.36	91.33	75.81
4.5 Mean weight	730.55	875.98	1,001.99	1,165.88	1,039.15

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	San Bernardino	Riverside	Hayward	Oakward	Remainder of Amameda
1. Children initial weights					
1.1 Number of sampled children	354	325	462	445	273
1.2 Sum of weights	341,565	299,993	66,081	73,447	100,941
1.3 Coefficient of variation (CV)	75.56	67.87	253.73	177.11	80.35
2. Nonresponse adjustment					
2.1 Number of completed child interviews	285	265	356	370	224
2.2 Sum of weights before adjustment	341,565	299,993	66,081	73,447	100,941
a. Eligible respondents	274,047	246,740	48,598	62,479	83,567
b. Ineligibles	189	3,235	53	397	0
c. Nonrespondents	67,328	50,019	17,429	10,571	17,374
2.3 Sum of weights after adjustment	341,565	299,993	66,081	73,447	100,941
a. Eligible respondents	341,347	295,982	66,020	72,978	100,941
b. ligibles	218	4,010	61	469	0
c. respondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	76.56	70.48	192.78	183.32	82.12
2.5 Mean adjustment factor	1.25	1.22	1.36	1.18	1.21
3. Trimming adjustment					
3.1 Number of trimmed records	3	2	3	1	2
3.2 Sum of weights before trimming adjustment	337,478	298,208	20,333	59,933	149,979
3.3 Sum of weights after trimming adjustment	334,159	295,852	18,408	56,182	149,176
4. Raking nonresponse adjustment					
4.1 Number of completed child interviews	282	268	191	345	392
4.2 Sum of weights after adjustment	357,430	322,236	27,427	65,641	148,886
4.3 Mean adjustment factor	1.07	1.09	1.49	1.17	1.00
4.4 Coefficient of variation (CV)	76.45	73.30	134.85	96.93	110.51
4.5 Mean weight	1,267.48	1,202.37	143.59	190.26	379.81

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	Sacramento	Contra Costa	Fresno	San Francisco	Ventura
1. Children initial weights					
1.1 Number of sampled children	251	203	212	141	146
1.2 Sum of weights	229,521	162,382	183,076	67,460	121,809
1.3 Coefficient of variation (CV)	82.12	65.76	73.97	61.06	80.60
2. Nonresponse adjustment					
2.1 Number of completed Child interviews	201	163	178	115	127
2.2 Sum of weights before adjustment	229,521	162,382	183,076	67,460	121,809
a. Eligible respondents	176,607	128,673	155,415	52,704	103,504
b. Ineligibles	2,404	935	2,771	1,036	5,086
c. Nonrespondents	50,510	32,775	24,890	13,721	13,219
2.3 Sum of weights after adjustment	229,521	162,382	183,076	67,460	121,809
a. Eligible respondents	226,210	161,146	179,540	66,139	116,012
b. Ineligibles	3,311	1,236	3,536	1,321	5,798
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	74.30	65.03	73.47	56.31	75.29
2.5 Mean adjustment factor	1.30	1.26	1.18	1.28	1.18
3. Trimming adjustment					
3.1 Number of trimmed records	3	2	0	1	0
3.2 Sum of weights before trimming adjustment	225,242	170,900	178,088	64,570	119,750
3.3 Sum of weights after trimming adjustment	221,863	169,678	178,088	64,158	119,750
4. Raking nonresponse adjustment					
4.1 Number of completed child interviews	200	185	177	112	132
4.2 Sum of weights after adjustment	226,958	167,699	163,219	82,592	138,624
4.3 Mean adjustment factor	1.02	0.99	0.92	1.29	1.16
4.4 Coefficient of variation (CV)	73.86	76.15	76.52	69.48	88.34
4.5 Mean weight	1,134.79	906.48	922.14	737.43	1,050.18

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	San Mateo	Kern	San Joaquin	Sonoma	Stanislaus
1. Children initial weights					
1.1 Number of sampled children	129	151	131	109	140
1.2 Sum of weights	101,217	145,727	90,500	66,039	82,201
1.3 Coefficient of variation (CV)	131.81	73.22	85.17	60.38	63.01
2. Nonresponse adjustment					
2.1 Number of completed child interviews	110	124	114	96	119
2.2 Sum of weights before adjustment	101,217	145,727	90,500	66,039	82,201
a. Eligible respondents	81,170	116,480	78,485	60,154	69,323
b. Ineligibles	564	0	0	0	369
c. Nonrespondents	19,483	29,247	12,016	5,886	12,509
2.3 Sum of weights after adjustment	101,217	145,727	90,500	66,039	82,201
a. Eligible respondents	100,584	145,727	90,500	66,039	81,777
b. Ineligibles	633	0	0	0	424
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	84.50	76.50	89.05	60.00	62.10
2.5 Mean adjustment factor	1.25	1.25	1.15	1.10	1.19
3. Trimming adjustment					
3.1 Number of trimmed records	1	2	2	1	1
3.2 Sum of weights before trimming adjustment	92,143	145,727	94,297	66,185	77,341
3.3 Sum of weights after trimming adjustment	90,895	144,522	88,562	65,715	76,700
4. Raking nonresponse adjustment					
4.1 Number of completed child interviews	107	124	115	97	114
4.2 Sum of weights after adjustment	110,806	143,612	120,371	69,481	92,529
4.3 Mean adjustment factor	1.22	0.99	1.36	1.06	1.21
4.4 Coefficient of variation (CV)	85.13	71.37	68.07	52.10	65.68
4.5 Mean weight	1,035.57	1,158.16	1,046.71	716.30	811.66

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	Santa Barbara	Solano	Tulare	Santa Cruz	Marin
1. Children initial weights					
1.1 Number of sampled children	127	146	178	108	106
1.2 Sum of weights	64,054	68,603	82,589	37,591	33,867
1.3 Coefficient of variation (CV)	76.36	63.65	75.80	64.64	56.49
2. Nonresponse adjustment					
2.1 Number of completed child interviews	107	113	142	87	94
2.2 Sum of weights before adjustment	64,054	68,603	82,589	37,591	33,867
a. Eligible respondents	54,817	50,223	63,196	30,159	29,915
b. Ineligibles	437	112	717	0	0
Nonrespondents	8,799	18,268	18,677	7,432	3,953
2.3 Sum of weights after adjustment	64,054	68,603	82,589	37,591	33,867
a a. Eligible respondents	63,576	68,462	81,678	37,591	33,867
b b. Ineligibles	478	141	911	0	0
c c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	78.91	60.03	68.58	66.35	57.84
2.5 Mean adjustment factor	1.17	1.37	1.31	1.25	1.13
3. Trimming adjustment					
3.1 Number of trimmed records	1	0	3	1	0
3.2 Sum of weights before trimming adjustment	60,062	65,018	83,787	33,393	33,867
3.3 Sum of weights after trimming adjustment	59,365	65,018	81,532	32,789	33,867
4. Raking nonresponse adjustment					
4.1 Number of completed child interviews	103	108	146	81	94
4.2 Sum of weights after adjustment	64,780	72,413	83,735	34,958	33,074
4.3 Mean adjustment factor	1.09	1.11	1.03	1.07	0.98
4.4 Coefficient of variation (CV)	70.58	65.22	63.62	61.84	63.06
4.5 Mean weight	628.93	670.49	573.53	431.58	351.85

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	San Luis Obispo	Placer	Merced	Butte	Shasta
1. Children initial weights					
1.1 Number of sampled children	96	120	170	107	97
1.2 Sum of weights	31,100	39,465	47,741	25,728	24,757
1.3 Coefficient of variation (CV)	71.29	91.96	90.03	76.42	60.48
2. Nonresponse adjustment					
2.1 Number of completed child interviews	83	97	141	98	81
2.2 Sum of weights before adjustment	31,100	39,465	47,741	25,728	24,757
a. Eligible respondents	27,184	31,335	37,870	23,968	21,512
b. Ineligibles	78	0	926	0	0
c. Nonrespondents	3,838	8,130	8,945	1,760	3,245
2.3 Sum of weights after adjustment	31,100	39,465	47,741	25,728	24,757
a. Eligible respondents	31,012	39,465	46,621	25,728	24,757
b. Ineligibles	88	0	1,120	0	0
Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	73.63	97.62	96.43	77.44	60.97
2.5 Mean adjustment factor	1.14	1.26	1.26	1.07	1.15
3. Trimming adjustment					
3.1 Number of trimmed records	1	3	1	1	1
3.2 Sum of weights before trimming adjustment	34,383	39,776	50,839	25,728	25,389
3.3 Sum of weights after trimming adjustment	33,811	36,614	49,941	25,484	25,204
4. Raking nonresponse adjustment					
4.1 Number of completed child interviews	86	99	145	98	85
4.2 Sum of weights after adjustment	32,886	44,985	48,688	28,277	26,197
4.3 Mean adjustment factor	0.97	1.23	0.97	1.11	1.04
4.4 Coefficient of variation (CV)	70.33	55.63	94.26	62.46	68.51
4.5 Mean weight	382.40	454.39	335.78	288.54	308.20

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	Yolo	El Dorado	Imperial	Napa	Kings
1. Children initial weights					
1.1 Number of sampled children	121	114	170	99	186
1.2 Sum of weights	25,299	23,909	25,227	16,385	25,940
1.3 Coefficient of variation (CV)	65.42	53.73	69.56	52.63	69.10
2. Nonresponse adjustment					
2.1 Number of completed child interviews	102	94	124	87	161
2.2 Sum of weights before adjustment	25,299	23,909	25,227	16,385	25,940
a. Eligible respondents	20,762	19,238	17,807	14,592	22,843
b. Ineligibles	0	323	518	0	55
c. Nonrespondents	4,537	4,348	6,901	1,793	3,042
2.3 Sum of weights after adjustment	25,299	23,909	25,227	16,385	25,940
a. Eligible respondents	25,299	23,508	24,507	16,385	25,884
b. Ineligibles	0	401	720	0	57
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	65.05	54.26	67.68	52.99	67.56
2.5 Mean adjustment factor	1.22	1.24	1.42	1.12	1.14
3. Trimming adjustment					
3.1 Number of trimmed records	0	0	0	1	1
3.2 Sum of weights before trimming adjustment	25,528	23,380	24,419	19,829	25,705
3.3 Sum of weights after trimming adjustment	25,528	23,380	24,419	19,184	25,575
4. Raking nonresponse adjustment					
4.1 Number of completed child interviews	104	93	123	92	160
4.2 Sum of weights after adjustment	28,747	24,302	28,179	17,249	26,031
4.3 Mean adjustment factor	1.13	1.04	1.15	0.90	1.02
4.4 Coefficient of variation (CV)	68.66	54.11	63.61	61.70	66.40
4.5 Mean weight	276.42	261.31	229.09	187.49	162.70

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake Mendocino
1. Children initial weights					
1.1 Number of sampled children	125	145	104	69	82
1.2 Sum of weights	19,769	83,458	24,894	14,968	23,159
1.3 Coefficient of variation (CV)	97.04	67.05	78.31	66.04	48.38
2. Nonresponse adjustment					
2.1 Number of completed child interviews	104	122	91	65	71
2.2 Sum of weights before adjustment	19,769	83,458	24,894	14,968	23,159
a. Eligible respondents	16,815	68,306	21,137	13,776	19,677
b. Ineligibles	0	0	0	0	672
c. Nonrespondents	2,954	15,152	3,757	1,192	2,810
2.3 Sum of weights after adjustment	19,769	83,458	24,894	14,968	23,159
a. Eligible respondents	19,769	83,458	24,894	14,968	22,355
b. Ineligibles	0	0	0	0	805
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	100.22	67.90	75.40	67.11	48.96
2.5 Mean adjustment factor	1.18	1.22	1.18	1.09	1.18
3. Trimming adjustment					
3.1 Number of trimmed records	3	2	1	0	0
3.2 Sum of weights before trimming adjustment	19,510	86,171	24,985	15,965	22,210
3.3 Sum of weights after trimming adjustment	17,706	84,841	24,135	15,965	22,210
4. Raking nonresponse adjustment					
4.1 Number of completed child interviews	103	126	92	68	70
4.2 Sum of weights after adjustment	25,209	89,180	23,009	11,473	21,981
4.3 Mean adjustment factor	1.42	1.05	0.95	0.72	0.99
4.4 Coefficient of variation (CV)	66.21	63.83	65.65	66.80	50.42
4.5 Mean weight	244.75	707.78	250.10	168.73	314.01

Table B-3. Extended interview weighting for child interview by stratum (combined RDD and surname list samples) (continued)

	Colusa, Glen, Tehama	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1. Children initial weights				
1.1 Number of sampled children	110	115	69	60
1.2 Sum of weights	19,641	25,837	14,365	25,442
1.3 Coefficient of variation (CV)	59.94	76.94	57.35	56.56
2. Nonresponse adjustment				
2.1 Number of completed child interviews	90	105	53	53
2.2 Sum of weights before adjustment	19,641	25,837	14,365	25,442
a. Eligible respondents	15,719	23,791	11,528	22,843
b. Ineligibles	0	0	311	0
c. Nonrespondents	3,922	2,046	2,526	2,599
2.3 Sum of weights after adjustment	19,641	25,837	14,365	25,442
a a. Eligible respondents	19,641	25,837	13,987	25,442
b b. Ineligibles	0	0	378	0
c c. Nonrespondents	0	0	0	0
2.4 Coefficient of variation (CV)	61.89	77.45	55.51	56.63
2.5 Mean adjustment factor	1.25	1.09	1.25	1.11
3. Trimming adjustment				
3.1 Number of trimmed records	2	2	1	0
3.2 Sum of weights before trimming adjustment	18,740	25,549	13,759	25,442
3.3 Sum of weights after trimming adjustment	18,012	24,725	13,313	25,442
4. Raking nonresponse adjustment				
4.1 Number of completed child interviews	85	102	49	53
4.2 Sum of weights after adjustment	18,809	27,628	12,206	21,914
4.3 Mean adjustment factor	1.04	1.12	0.92	0.86
4.4 Coefficient of variation (CV)	62.44	75.49	58.73	60.02
4.5 Mean weight	221.28	270.86	249.11	413.48

Table B-4. 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	6,857	1,671	347	389	210
1.2 Sum of weights	3,188,004	853,726	264,617	266,943	135,641
1.3 Coefficient of variation (CV)	100.03	81.08	87.21	94.64	74.19
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	4,010	925	208	201	123
2.2 Sum of weights before adjustment	3,188,004	853,726	264,617	266,943	135,641
a. Eligible respondents	1,796,144	471,191	155,929	130,315	78,626
b. Ineligibles	51,376	19,239	3,673	1,585	4,663
c. Nonrespondents	1,340,484	363,296	105,015	135,042	52,352
2.3 Sum of weights after adjustment	3,188,003	853,726	264,617	266,943	135,641
a. Eligible respondents	3,100,870	820,141	258,810	263,743	128,115
b. Ineligibles	87,133	33,585	5,807	3,200	7,526
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	102.66	81.14	87.35	95.47	77.18
2.5 Mean adjustment factor	1.77	1.81	1.70	2.05	1.73
3. Trimming adjustment					
3.1 Number of trimmed records	63	2	2	0	2
3.2 Sum of weights before trimming adjustment	3,100,870	826,638	258,810	252,020	133,712
3.3 Sum of weights after trimming adjustment	3,053,242	824,549	254,222	252,020	132,507
4. Raking nonresponse adjustment					
4.1 Number of completed adolescent interviews	4,010	928	208	196	126
4.2 Sum of weights after adjustment	3,259,771	899,796	257,576	259,379	134,872
4.3 Mean adjustment factor	1.07	1.09	1.01	1.03	1.02
4.4 Coefficient of variation (CV)	103.24	86.51	85.80	97.18	83.14
4.5 Mean weight	812.91	969.61	1,238.35	1,323.36	1,070.41

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

	San Bernardino	Riverside	Hayward	Oakward	Remainder of Amameda
1. Adolescent initial weights					
1.1 Number of sampled adolescents	268	230	270	235	173
1.2 Sum of weights	202,929	189,051	27,029	28,967	54,893
1.3 Coefficient of variation (CV)	67.82	85.06	201.17	115.15	74.15
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	149	136	158	144	101
2.2 Sum of weights before adjustment	202,929	189,051	27,029	28,967	54,893
a. Eligible respondents	112,202	103,553	13,900	18,241	28,553
b. Ineligibles	537	1,479	671	445	1,699
c. Nonrespondents	90,191	84,020	12,459	10,282	24,641
2.3 Sum of weights after adjustment	202,929	189,051	27,029	28,967	54,893
a. Eligible respondents	201,867	186,492	25,845	28,220	51,641
b. Ineligibles	1,063	2,560	1,185	748	3,252
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	67.10	78.78	171.82	134.83	74.88
2.5 Mean adjustment factor	1.81	1.83	1.94	1.59	1.92
3. Trimming adjustment					
3.1 Number of trimmed records	1	3	1	1	2
3.2 Sum of weights before trimming adjustment	202,483	186,492	8,362	24,179	70,528
3.3 Sum of weights after trimming adjustment	201,424	180,082	7,906	24,050	69,594
4. Raking nonresponse adjustment					
4.1 Number of completed adolescent interviews	150	136	78	139	180
4.2 Sum of weights after adjustment	211,384	182,180	13,002	35,939	67,574
4.3 Mean adjustment factor	1.05	1.01	1.64	1.49	0.97
4.4 Coefficient of variation (CV)	70.61	77.66	106.33	113.35	106.68
4.5 Mean weight	1,409.23	1,339.56	166.69	258.56	375.41

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

	Sacramento	Contra Costa	Fresno	San Francisco	Ventura
1. Adolescent initial weights					
1.1 Number of sampled adolescents	152	137	124	59	98
1.2 Sum of weights	116,367	90,029	82,746	24,813	66,314
1.3 Coefficient of variation (CV)	81.72	65.88	65.67	82.18	54.43
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	81	87	66	36	59
2.2 Sum of weights before adjustment	116,367	90,029	82,746	24,813	66,314
a. Eligible respondents	59,914	58,298	47,185	14,382	39,992
b. Ineligibles	3,965	0	693	0	558
c. Nonrespondents	52,487	31,731	34,868	10,431	25,764
2.3 Sum of weights after adjustment	116,367	90,029	82,746	24,813	66,314
a. Eligible respondents	109,913	90,029	81,619	24,813	65,402
b. Ineligibles	6,454	0	1,127	0	912
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	67.21	64.75	64.99	58.18	54.05
2.5 Mean adjustment factor	1.94	1.54	1.75	1.73	1.66
3. Trimming adjustment					
3.1 Number of trimmed records	1	1	2	3	0
3.2 Sum of weights before trimming adjustment	109,913	93,187	80,861	23,920	65,402
3.3 Sum of weights after trimming adjustment	109,791	92,384	78,761	22,894	65,402
4. Raking nonresponse adjustment					
4.1 Number of completed adolescent interviews	81	94	66	35	59
4.2 Sum of weights after adjustment	124,941	92,250	93,809	34,248	77,096
4.3 Mean adjustment factor	1.14	1.00	1.19	1.50	1.18
4.4 Coefficient of variation (CV)	70.28	75.97	80.13	68.36	65.83
4.5 Mean weight	1,542.48	981.38	1,421.35	978.51	1,306.72

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

	San Mateo	Kern	San Joaquin	Sonoma	Stanislaus
1. Adolescent initial weights					
1.1 Number of sampled adolescents	99	111	105	71	112
1.2 Sum of weights	65,928	79,695	62,665	38,238	58,979
1.3 Coefficient of variation (CV)	80.80	62.65	71.14	68.48	73.64
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	54	64	62	39	65
2.2 Sum of weights before adjustment	65,928	79,695	62,665	38,238	58,979
a. Eligible respondents	33,687	44,637	32,043	21,325	35,918
b. Ineligibles	0	2,815	1,445	619	0
c. Nonrespondents	32,241	32,243	29,176	16,295	23,061
2.3 Sum of weights after adjustment	65,928	79,695	62,665	38,238	58,979
a. Eligible respondents	65,928	74,936	59,961	37,160	58,979
b. ligibles	0	4,759	2,704	1,078	0
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	85.52	65.34	73.84	78.76	77.56
2.5 Mean adjustment factor	1.96	1.79	1.96	1.79	1.64
3. Trimming adjustment					
3.1 Number of trimmed records	3	2	3	1	0
3.2 Sum of weights before trimming adjustment	66,009	75,066	60,231	37,160	57,654
3.3 Sum of weights after trimming adjustment	60,256	74,005	55,996	35,500	57,654
4. Raking nonresponse adjustment					
4.1 Number of completed adolescent interviews	54	65	63	39	63
4.2 Sum of weights after adjustment	52,808	74,706	66,920	39,832	52,430
4.3 Mean adjustment factor	0.88	1.01	1.20	1.12	0.91
4.4 Coefficient of variation (CV)	70.78	61.25	58.39	70.28	74.92
4.5 Mean weight	977.92	1,149.32	1,062.21	1,021.34	832.23

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

	Santa Barbara	Solano	Tulare	Santa Cruz	Marin
1. Adolescent initial weights					
1.1 Number of sampled adolescents	90	113	127	71	50
1.2 Sum of weights	34,213	46,594	45,025	20,075	11,935
1.3 Coefficient of variation (CV)	60.60	68.79	66.09	57.85	55.90
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	59	67	82	48	31
2.2 Sum of weights before adjustment	34,213	46,594	45,025	20,075	11,935
a. Eligible respondents	22,747	27,893	28,089	13,773	6,868
b. Ineligibles	437	333	0	0	175
c. Nonrespondents	11,028	18,368	16,936	6,302	4,891
2.3 Sum of weights after adjustment	34,213	46,594	45,025	20,075	11,935
a. Eligible respondents	33,568	46,049	45,025	20,075	11,638
b. Ineligibles	645	545	0	0	297
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	64.88	59.60	66.76	61.70	36.14
2.5 Mean adjustment factor	1.50	1.67	1.60	1.46	1.74
3. Trimming adjustment					
3.1 Number of trimmed records	3	3	3	1	2
3.2 Sum of weights before trimming adjustment	33,568	44,971	45,308	20,075	11,638
3.3 Sum of weights after trimming adjustment	31,720	43,828	44,459	19,718	11,269
4. Raking nonresponse adjustment					
4.1 Number of completed adolescent interviews	59	66	82	48	31
4.2 Sum of weights after adjustment	34,714	38,998	44,209	24,668	16,933
4.3 Mean adjustment factor	1.09	0.89	0.99	1.25	1.50
4.4 Coefficient of variation (CV)	58.66	66.81	72.14	69.99	43.00
4.5 Mean weight	588.37	590.89	539.13	513.91	546.22

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

	San Luis Obispo	Placer	Merced	Butte	Shasta
1. Adolescent initial weights					
1.1 Number of sampled adolescents	75	92	110	77	75
1.2 Sum of weights	22,100	23,585	24,037	17,158	16,548
1.3 Coefficient of variation (CV)	64.89	47.84	60.06	71.95	54.19
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	46	56	69	53	43
2.2 Sum of weights before adjustment	22,100	23,585	24,037	17,158	16,548
a. Eligible respondents	13,923	15,651	15,085	10,411	9,014
b. Ineligibles	0	237	748	0	0
c. Nonrespondents	8,177	7,697	8,204	6,748	7,534
2.3 Sum of weights after adjustment	22,100	23,585	24,037	17,158	16,548
a. Eligible respondents	22,100	23,233	22,901	17,158	16,548
b. Ineligibles	0	352	1,136	0	0
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	68.08	49.53	61.80	56.89	62.38
2.5 Mean adjustment factor	1.59	1.51	1.59	1.65	1.84
3. Trimming adjustment					
3.1 Number of trimmed records	3	0	2	1	0
3.2 Sum of weights before trimming adjustment	22,100	22,990	24,652	17,870	18,360
3.3 Sum of weights after trimming adjustment	20,755	22,990	24,294	17,672	18,360
4. Raking nonresponse adjustment					
4.1 Number of completed adolescent interviews	46	57	72	54	49
4.2 Sum of weights after adjustment	18,958	27,312	26,924	19,042	15,408
4.3 Mean adjustment factor	0.91	1.19	1.11	1.08	0.84
4.4 Coefficient of variation (CV)	73.00	68.55	61.95	52.89	63.09
4.5 Mean weight	412.13	479.16	373.94	352.63	314.45

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

	Yolo	El Dorado	Imperial	Napa	Kings
1. Adolescent initial weights					
1.1 Number of sampled adolescents	90	91	138	67	122
1.2 Sum of weights	16,091	17,914	15,790	11,374	15,239
1.3 Coefficient of variation (CV)	61.82	60.07	50.51	82.76	83.45
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	56	55	85	43	72
2.2 Sum of weights before adjustment	16,091	17,914	15,790	11,374	15,239
a. Eligible respondents	9,384	10,020	10,402	7,792	9,658
b. Ineligibles	102	606	133	0	249
c. Nonrespondents	6,605	7,288	5,255	3,582	5,331
2.3 Sum of weights after adjustment	16,091	17,914	15,790	11,374	15,239
a. Eligible respondents	15,918	16,892	15,574	11,374	14,870
b. Ineligibles	173	1,021	215	0	368
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	59.18	59.72	54.32	83.50	81.11
2.5 Mean adjustment factor	1.71	1.79	1.52	1.46	1.58
3. Trimming adjustment					
3.1 Number of trimmed records	1	1	0	1	1
3.2 Sum of weights before trimming adjustment	15,929	17,412	15,574	12,704	14,781
3.3 Sum of weights after trimming adjustment	15,652	17,278	15,574	12,421	14,433
4. Raking nonresponse adjustment					
4.1 Number of completed adolescent interviews	56	55	85	45	71
4.2 Sum of weights after adjustment	16,392	17,086	17,425	13,136	12,995
4.3 Mean adjustment factor	1.05	0.99	1.12	1.06	0.90
4.4 Coefficient of variation (CV)	57.16	59.52	59.59	78.88	82.32
4.5 Mean weight	292.71	310.66	205.00	291.92	183.03

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

	Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake Mendocino
1. Adolescent initial weights					
1.1 Number of sampled adolescents	100	83	68	55	54
1.2 Sum of weights	13,121	38,952	13,745	8,248	11,534
1.3 Coefficient of variation (CV)	52.05	67.67	82.19	63.84	52.94
2. Nonresponse adjustment					
2.1 Number of completed adolescent interviews	68	44	44	37	32
2.2 Sum of weights before adjustment	13,121	38,952	13,745	8,248	11,534
a. Eligible respondents	8,997	21,532	8,214	5,747	7,153
b. Ineligibles	0	481	252	0	67
c. Nonrespondents	4,124	16,940	5,279	2,501	4,313
2.3 Sum of weights after adjustment	13,121	38,952	13,745	8,248	11,534
a. Eligible respondents	13,121	38,102	13,335	8,248	11,426
b. Ineligibles	0	850	409	0	107
c. Nonrespondents	0	0	0	0	0
2.4 Coefficient of variation (CV)	58.43	76.18	59.43	65.19	52.46
2.5 Mean adjustment factor	1.46	1.81	1.67	1.44	1.61
3. Trimming adjustment					
3.1 Number of trimmed records	0	3	1	0	1
3.2 Sum of weights before trimming adjustment	13,034	38,102	12,778	9,268	11,426
3.3 Sum of weights after trimming adjustment	13,034	35,521	12,358	9,268	11,292
4. Raking nonresponse adjustment					
4.1 Number of completed adolescent interviews	67	44	43	38	32
4.2 Sum of weights after adjustment	12,593	42,233	12,116	9,811	13,951
4.3 Mean adjustment factor	0.97	1.19	0.98	1.06	1.24
4.4 Coefficient of variation (CV)	62.16	57.95	59.36	60.39	49.48
4.5 Mean weight	187.96	959.84	281.76	258.17	435.98

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

	Colusa, Glen, Tehama	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1. Adolescent initial weights				
1.1 Number of sampled adolescents	69	76	54	49
1.2 Sum of weights	10,262	14,957	10,089	19,849
1.3 Coefficient of variation (CV)	71.07	81.22	63.76	72.69
2. Nonresponse adjustment				
2.1 Number of completed adolescent interviews	41	52	38	31
2.2 Sum of weights before adjustment	10,262	14,957	10,089	19,849
a. Eligible respondents	5,849	10,458	7,125	10,468
b. Ineligibles	0	188	194	3,089
c. Nonrespondents	4,413	4,311	2,770	6,292
2.3 Sum of weights after adjustment	10,262	14,957	10,089	19,849
a. Eligible respondents	10,262	14,692	9,822	15,326
b. Ineligibles	0	264	267	4,523
c. Nonrespondents	0	0	0	0
2.4 Coefficient of variation (CV)	77.30	91.13	61.87	72.86
2.5 Mean adjustment factor	1.75	1.43	1.42	1.90
3. Trimming adjustment				
3.1 Number of trimmed records	0	0	4	2
3.2 Sum of weights before trimming adjustment	8,450	13,267	9,359	14,629
3.3 Sum of weights after trimming adjustment	8,450	13,267	8,131	12,503
4. Raking nonresponse adjustment				
4.1 Number of completed adolescent interviews	35	48	38	29
4.2 Sum of weights after adjustment	9,176	15,563	13,334	16,054
4.3 Mean adjustment factor	1.09	1.17	1.64	1.28
4.4 Coefficient of variation (CV)	68.61	77.70	44.28	46.78
4.5 Mean weight	262.19	324.24	350.89	553.57

Appendix C

Table C-1. Average *DEFF* and *DEFT* for estimates from the adult interview for continuous and categorical variables

County/strata	Categorical variables					Continuous variables				
	Design effect (*DEFF*)				*DEFT*	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
State Total	1.65	1.73	2.43	0.53	1.28	1.80	1.43	2.93	0.87	1.34
Los Angeles	1.34	1.31	1.98	0.82	1.16	1.35	1.28	2.17	0.82	1.16
San Diego	1.55	1.60	2.21	1.03	1.24	1.47	1.48	2.09	0.88	1.21
Orange	1.56	1.49	2.34	0.94	1.25	1.49	1.43	1.94	1.13	1.22
Santa Clara	1.28	1.28	1.73	0.69	1.13	1.13	1.08	1.44	0.94	1.06
San Bernardino	1.41	1.35	2.10	0.98	1.19	1.00	0.96	1.18	0.83	1.00
Riverside	1.33	1.29	1.79	0.94	1.15	1.34	1.43	1.58	0.93	1.16
Alameda	2.49	2.36	4.23	1.69	1.58	2.45	2.41	3.39	1.61	1.57
Sacramento	1.41	1.38	2.40	0.74	1.19	1.30	1.32	1.68	1.05	1.14
Contra Costa	1.34	1.30	1.98	0.63	1.16	1.18	1.10	2.39	0.63	1.09
Fresno	1.37	1.40	1.83	0.79	1.17	1.14	0.90	2.10	0.67	1.07
San Francisco	1.26	1.24	1.83	0.63	1.12	1.22	1.30	1.35	0.97	1.10
Ventura	1.54	1.56	2.89	0.44	1.24	1.22	1.13	1.84	0.73	1.10
San Mateo	1.34	1.33	2.14	0.62	1.16	1.21	1.19	1.65	0.88	1.10
Kern	1.33	1.26	1.94	0.84	1.15	1.37	1.53	2.01	0.71	1.17
San Joaquin	1.37	1.36	2.06	0.50	1.17	1.18	1.25	1.51	0.77	1.09
Sonoma	1.29	1.31	1.78	0.88	1.14	1.21	1.16	1.87	0.76	1.10
Stanislaus	1.30	1.23	2.13	0.71	1.14	1.14	1.16	1.69	0.67	1.07
Santa Barbara	1.41	1.42	2.18	0.65	1.19	1.21	1.02	2.49	0.65	1.10
Solano	1.49	1.50	2.54	0.93	1.22	1.20	1.13	1.88	0.72	1.10
Tulare	1.24	1.22	1.84	0.78	1.12	1.07	1.18	1.26	0.79	1.04
Santa Cruz	1.34	1.36	1.90	0.86	1.16	1.09	0.94	1.75	0.76	1.04
Marin	1.31	1.35	2.24	0.52	1.14	1.17	1.10	1.52	0.83	1.08
San Luis Obispo	1.22	1.19	1.71	0.73	1.10	1.21	1.01	1.80	0.73	1.10
Placer	1.34	1.26	2.32	0.67	1.16	1.41	1.52	1.82	0.84	1.19
Merced	1.41	1.37	2.47	0.80	1.19	1.38	1.25	1.89	0.82	1.18
Butte	1.24	1.20	1.85	0.75	1.12	1.24	1.23	1.54	0.82	1.11

Table C-1. Average *DEFF* and *DEFT* for estimates from the adult interview for continuous and categorical variables (continued)

County/strata	Categorical variables					Continuous variables				
	Design effect (*DEFF*)				*DEFT*	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
Shasta	1.28	1.28	2.19	0.74	1.13	1.08	1.18	1.29	0.85	1.04
Yolo	1.35	1.35	1.86	0.63	1.16	1.11	0.97	1.63	0.87	1.05
El Dorado	1.20	1.18	1.72	0.50	1.10	1.13	0.87	1.81	0.84	1.06
Imperial	1.37	1.33	2.05	0.73	1.17	1.10	1.15	1.37	0.75	1.05
Napa	1.32	1.37	2.03	0.26	1.15	1.59	1.15	3.82	0.73	1.26
Kings	1.27	1.25	1.76	0.36	1.13	1.47	1.48	1.95	1.06	1.21
Madera	1.29	1.35	1.70	0.66	1.14	1.15	1.02	1.83	0.82	1.07
Monterey, San Benito	1.22	1.29	1.59	0.75	1.10	1.17	1.00	2.01	0.82	1.08
Del Norte, Humboldt	1.15	1.14	1.55	0.68	1.07	1.23	1.20	1.74	0.93	1.11
Lassen, Modoc, Siskiyou, Trinity	1.20	1.19	1.92	0.60	1.10	1.23	1.24	1.48	0.84	1.11
Lake, Mendocino	1.28	1.21	1.77	0.85	1.13	1.10	1.06	1.52	0.74	1.05
Colusa, Glen, Tehama	1.15	1.09	2.80	0.49	1.07	1.26	1.29	1.66	0.90	1.12
Sutter, Yuba	1.18	1.18	1.67	0.48	1.09	1.25	1.20	2.01	0.73	1.12
Plumas, Nevada, Sierra	1.35	1.34	2.09	0.75	1.16	1.41	1.31	2.41	0.63	1.19
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.23	1.24	1.89	0.76	1.11	1.18	1.03	1.63	0.84	1.09

Table C-2. Average *DEFF* and *DEFT* for estimates from the child interview for continuous and categorical variables

County/strata	Categorical variables					Continuous variables				
	Design effect (*DEFF*)				*DEFT*	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
State Total	2.00	1.95	3.68	1.16	1.41	2.21	2.43	2.82	1.54	1.49
Los Angeles	1.85	1.79	2.97	1.39	1.36	1.96	2.07	2.32	1.47	1.40
San Diego	1.62	1.70	2.54	0.71	1.18	1.53	1.54	1.99	1.23	1.24
Orange	1.90	1.81	3.12	1.14	1.18	1.70	1.59	2.27	1.22	1.30
Santa Clara	1.79	1.87	338	0.27	1.16	2.25	1.89	4.92	0.65	1.50
San Bernardino	1.33	1.27	2.39	0.88	1.26	1.37	1.29	1.81	0.93	1.17
Riverside	1.50	1.33	3.91	1.01	1.19	1.51	1.23	2.14	1.02	1.23
Alameda	1.93	2.18	2.67	0.39	1.42	2.27	2.11	3.24	1.66	1.51
Sacramento	1.19	1.13	2.38	0.47	1.23	1.64	1.38	2.36	1.28	1.28
Contra Costa	1.29	1.29	1.85	0.63	1.06	1.18	1.18	1.74	0.76	1.09
Fresno	1.42	1.19	2.55	0.84	1.25	1.37	1.04	2.41	0.90	1.17
San Francisco	1.13	1.02	1.95	0.76	1.28	1.23	1.35	1.74	0.52	1.11
Ventura	1.47	1.48	1.94	1.05	1.17	1.22	0.97	2.68	0.51	1.10
San Mateo	1.57	1.66	2.62	0.00	1.10	1.54	1.33	3.28	0.57	1.24
Kern	1.34	1.31	1.82	0.89	1.23	1.25	1.20	1.64	0.91	1.12
San Joaquin	1.17	1.10	1.81	0.66	1.18	0.93	0.90	1.66	0.51	0.97
Sonoma	1.21	1.34	1.65	0.53	1.29	1.38	1.23	2.52	0.58	1.18
Stanislaus	1.70	1.71	2.74	0.50	1.13	1.48	1.43	2.22	0.37	1.22
Santa Barbara	1.25	1.40	1.59	0.65	1.18	1.21	1.29	1.38	0.88	1.10
Solano	1.30	1.32	1.76	0.68	1.19	1.77	1.17	3.90	0.75	1.33
Tulare	1.47	1.43	2.18	0.39	1.27	1.05	0.92	1.46	0.58	1.02
Santa Cruz	1.23	1.19	1.91	0.61	1.24	1.05	1.04	1.68	0.57	1.02
Marin	1.00	0.97	1.71	0.61	1.10	1.14	0.87	2.06	0.64	1.07
San Luis Obispo	1.22	1.26	1.73	0.57	1.09	1.46	1.38	2.28	0.95	1.21
Placer	1.10	1.12	1.46	0.40	1.42	1.20	1.25	1.74	0.75	1.10
Merced	1.97	1.43	2.12	0.81	1.34	1.31	1.28	1.88	0.81	1.14
Butte	1.16	1.15	1.45	0.79	1.07	1.58	1.48	2.07	1.15	1.26
Shasta	1.26	1.16	1.88	0.82	1.26	1.30	1.29	2.19	0.73	1.14
Yolo	1.15	1.21	1.59	0.39	1.07	1.32	1.21	1.92	0.99	1.15

Table C-2. Average *DEFF* and *DEFT* for estimates from the child interview for continuous and categorical variables (continued)

County/strata	Categorical variables					Continuous variables				
	Design effect (*DEFF*)				*DEFT*	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
El Dorado	1.17	1.21	2.07	0.50	1.29	1.08	1.14	1.48	0.59	1.04
Imperial	1.41	1.38	1.98	0.98	1.13	1.52	1.22	2.85	1.02	1.23
Napa	1.80	1.91	2.77	0.90	1.26	1.14	1.06	1.77	0.38	1.07
Kings	2.05	2.04	3.96	0.49	1.29	2.03	1.75	3.40	1.19	1.42
Madera	1.52	1.50	2.31	0.81	1.19	1.87	1.88	3.31	0.95	1.37
Monterey, San Benito	1.19	1.25	1.72	0.60	1.20	1.48	1.41	2.24	0.91	1.22
Del Norte, Humboldt	1.10	1.05	1.70	0.56	1.14	1.07	1.05	1.37	0.81	1.03
Lassen, Modoc, Siskiyou, Trinity	1.42	1.40	2.06	0.75	1.19	1.76	1.07	4.57	0.80	1.33
Lake, Mendocino	1.24	1.24	1.65	0.87	1.30	1.06	0.97	1.47	0.92	1.03
Colusa, Glen, Tehama	1.41	1.43	2.16	0.80	1.24	1.68	1.77	2.24	1.29	1.30
Sutter, Yuba	1.68	1.71	3.85	0.69	1.16	1.29	0.94	3.08	0.39	1.13
Plumas, Nevada, Sierra	1.28	1.13	2.28	0.83	1.12	1.44	1.37	1.82	1.24	1.20
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.14	1.08	2.11	0.79	1.25	0.99	0.97	1.20	0.89	1.00

Table C-3. Average *DEFF* and *DEFT* for estimates from the adolescent interview for continuous and categorical variables

County/strata	Categorical variables					Continuous variables				
	Design effect (*DEFF*)				*DEFT*	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
State Total	2.03	1.95	2.82	1.16	1.43	1.91	1.77	2.44	1.37	1.38
Los Angeles	1.86	1.79	2.97	1.39	1.36	1.78	1.84	2.32	1.17	1.33
San Diego	1.62	1.70	2.54	0.71	1.27	1.46	1.48	1.62	1.28	1.21
Orange	1.90	1.81	3.12	1.14	1.38	1.70	1.56	2.32	1.22	1.30
Santa Clara	1.83	1.89	3.38	0.27	1.35	1.93	1.33	4.92	0.65	1.39
San Bernardino	1.37	1.27	2.39	0.88	1.17	1.21	1.23	1.53	0.93	1.10
Riverside	1.45	1.24	3.91	1.01	1.21	1.50	1.34	2.14	0.71	1.22
Alameda	1.93	2.13	2.67	0.39	1.39	2.20	2.06	3.24	1.35	1.45
Sacramento	1.20	1.16	2.38	0.47	1.10	1.65	1.58	2.36	1.12	1.29
Contra Costa	1.31	1.30	1.85	0.63	1.15	1.23	1.18	1.97	0.76	1.11
Fresno	1.40	1.19	2.55	0.84	1.18	1.40	1.04	2.41	0.85	1.19
San Francisco	1.16	1.05	1.95	0.76	1.08	1.30	1.37	1.79	0.52	1.14
Ventura	1.43	1.45	1.94	0.97	1.20	1.14	0.65	2.68	0.44	1.07
San Mateo	1.54	1.66	2.62	0.00	1.24	1.86	1.33	3.28	0.57	1.36
Kern	1.32	1.31	1.82	0.89	1.15	1.23	1.20	1.64	0.74	1.11
San Joaquin	1.13	1.04	1.81	0.66	1.06	1.05	0.90	1.66	0.51	1.02
Sonoma	1.18	1.33	1.65	0.53	1.09	1.38	1.13	2.52	0.73	1.18
Stanislaus	1.65	1.71	2.74	0.37	1.29	1.33	1.32	2.22	0.73	1.15
Santa Barbara	1.27	1.39	1.59	0.65	1.12	0.95	1.04	1.29	0.40	0.98
Solano	1.33	1.32	2.19	0.68	1.15	1.76	1.47	3.90	0.75	1.33
Tulare	1.44	1.41	2.18	0.39	1.20	1.24	1.46	1.76	0.58	1.11
Santa Cruz	1.19	1.14	1.91	0.61	1.09	1.14	1.22	1.68	0.57	1.07
Marin	0.97	0.95	1.71	0.61	0.98	1.30	1.11	2.06	0.87	1.14
San Luis Obispo	1.30	1.28	2.28	0.57	1.14	1.30	1.36	1.82	0.95	1.14
Placer	1.12	1.12	1.74	0.40	1.06	1.15	1.24	1.29	0.75	1.07
Merced	1.48	1.43	2.12	0.81	1.22	1.40	1.12	2.53	0.81	1.19
Butte	1.17	1.15	1.45	0.79	1.08	1.66	1.71	2.07	1.19	1.29
Shasta	1.32	1.17	2.19	0.82	1.15	1.08	0.96	1.48	0.73	1.04
Yolo	1.16	1.21	1.59	0.39	1.08	1.23	1.21	1.92	0.62	1.11

Table C-3. Average *DEFF* and *DEFT* for estimates from the adolescent interview for continuous and categorical variables (continued)

County/strata	Categorical variables					Continuous variables				
	Design effect (*DEFF*)				*DEFT*	Design effect (*DEFF*)				*DEFT*
	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
El Dorado	1.15	1.21	2.07	0.50	1.07	1.09	0.99	1.48	0.84	1.04
Imperial	1.40	1.38	1.98	0.98	1.18	1.50	1.22	2.85	1.05	1.22
Napa	1.72	1.77	2.77	0.38	1.31	1.38	1.59	1.74	0.89	1.17
Kings	2.08	2.04	3.96	0.49	1.44	1.67	1.53	2.44	1.19	1.29
Madera	1.57	1.57	2.31	0.81	1.25	1.63	1.33	3.31	0.95	1.28
Monterey, San Benito	1.21	1.26	1.72	0.60	1.10	1.46	1.34	2.24	0.91	1.21
Del Norte, Humboldt	1.09	1.03	1.70	0.56	1.05	1.07	1.08	1.37	0.81	1.04
Lassen, Modoc, Siskiyou, Trinity	1.39	1.40	2.06	0.75	1.18	1.97	1.12	4.57	0.90	1.40
Lake, Mendocino	1.20	1.19	1.65	0.87	1.10	1.17	1.11	1.47	0.95	1.08
Colusa, Glen, Tehama	1.40	1.33	2.16	0.80	1.18	1.56	1.77	2.24	0.96	1.25
Sutter, Yuba	1.57	1.69	3.85	0.39	1.25	1.77	1.51	3.08	0.94	1.33
Plumas, Nevada, Sierra	1.28	1.14	2.28	0.83	1.13	1.33	1.24	1.82	1.01	1.15
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.11	1.05	2.11	0.79	1.05	1.05	1.00	1.20	0.88	1.02