California
Health
Interview
Survey

Making California's Voices Heard on Health

CHIS 2001 Methodology Series

Report 5

Weighting and Variance Estimation

CALIFORNIA HEALTH INTERVIEW SURVEY

CHIS 2001 METHODOLOGY SERIES

REPORT 5

WEIGHTING AND VARIANCE ESTIMATION

September 2002 *(rev 1/29/03)*

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

www.chis.ucla.edu

This report describes the weighting and variance estimation methods used in CHIS 2001. This report presents the steps used to create the analytical weights for analyzing the data from the adult, child, adolescent, adolescent insurance interviews.

Suggested citation:

California Health Interview Survey. *CHIS 2001 Methodology Series: Report 5 – Weighting and Variance Estimation.* Los Angeles, CA: UCLA Center for Health Policy Research, 2002.

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 2001 came from multiple sources: the California Department of Health Services, The California Endowment, the California Children and Families Commission, the National Cancer Institute, the Centers for Disease Control and Prevention, and the Indian Health Service.

PREFACE

Weighting and Variance Estimation in CHIS 2001 is the fifth in a series of methodological reports describing the 2001 California Health Interview Survey (CHIS 2001). 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 2001 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 the 2001 CHIS are as follows:

- Report 1: Sample Design for CHIS 2001;
- Report 2: Data Collection Methods in CHIS 2001;
- Report 3: Data Processing Procedures in CHIS 2001;
- Report 4: Response Rates in CHIS 2001; and
- Report 5: Weighting and Variance Estimation for CHIS 2001.

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 2001. 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, adolescent, and adolescent insurance interviews.

TABLE OF CONTENTS

Chapter			Page
1	CHIS 2001 DESIGN AND METHODOLOGY SUMMARY		1-1
	1.1	Overview	1-1
	1.2	Sample Design Objectives	1-2
	1.3	Data Collection	1-3
	1.4	Response Rate	1-5
	1.5	Weighting the Random Digit Dial Sample	1-6
	1.6	Imputation Methods	1-8
	1.7	Methodology Report Series	1-9
2	WEIGHTING ADJUSTMENTS FOR RANDOM-DIGIT-DIALING AND GEOGRAPHIC SAMPLE WEIGHTS		2-1
	2.1	Weighting Approach	2-1
	2.2	Weight Adjustment Method	2-4
3	HOUSEHOLD WEIGHTING		3-1
	3.1	Base Weights	3-1
	3.2	Subsampling Adjustments	3-3
	3.3	Unknown Residential Status Adjustment	3-7
	3.4	Screener Nonresponse Adjustment	3-9
	3.5	Multiple Telephone Adjustment	3-10
	3.6	Household Poststratification	3-11
4	ADULT WEIGHTING		4-1
	4.1	Adult Initial Weight	4-1
	4.2	Adult Nonresponse Adjustment	4-2
	4.3	Adult Trimming Factors	4-4
	4.4	Adult Raked Weight	4-5
	4.5	Nontelephone Adjustments	4-6
5	CHILD WEIGHTING		5-1
	5.1	Household-Level Adjustment	5-1
	5.2	Initial Child Weighting	5-2
	5.3	Other Child Weighting Adjustments	5-3
6	ADOLESCENT WEIGHTING		6-1
	6.1	Initial Adolescent and Adolescent Insurance Weights	6-1
	6.2	Other Adolescent Weighting Adjustments	6-2

TABLE OF CONTENTS (continued)

Chapter			Page
7	RAKING AND CONTROL TOTALS		7-1
	7.1	Raking Procedure	7-1
	7.2	Raking Dimensions	7-2
	7.3	Producing the Control Totals	7-7
8	IMPUTATION PROCEDURES		8-1
	8.1	Imputed Variables and Methods	8-1
	8.2	Self-Reported County and Self-Reported Stratum	8-3
	8.3	Self-Reported Sex and Age	8-7
	8.4	Self-Reported Race and Ethnicity	8-8
9	WEIGHTING ADJUSTMENTS FOR THE RACE-ETHNIC SUPPLEMENTAL SAMPLES		9-1
	9.1	Sampling Options	9-2
	9.2	Model-based and Design-based Approaches to Weighting	9-3
	9.3	Creation of Base Weights	9-5
	9.4	Adjustment Factors	9-8
	9.5	Raking Adjustment	9-11
	9.6	Evaluation of the Weights	9-20
10	VARIANCE ESTIMATION		10-1
	10.1	Design Effects	10-1
	10.2	Methods for Variance Estimation	10-9
	10.3	Design of Replicates	10-10
	10.4	Software for Computing Variances	10-12
	REFERENCES		R-1

List of Tables

Table		
1-1	California county and county group strata used in the sample design	1-2
1-2	Number of completed interviews by type of sample, instrument	1-4
1-3	Survey topic areas by instrument	1-5

TABLE OF CONTENTS (continued)

List of Tables (continued)

Table		Page
2-1	CHIS 2001 weights and type of estimate	2-2
3-1	Number of banks for CHIS 2001 sample selection, average number of banks, sample size, and base weight by sampling strata	3-4
3-2	Sample sizes for the San Francisco County and Santa Barbara County supplemental samples	3-5
3-3	Substratum definition for subsampling and planned subsampling rate	3-6
3-4	Estimated residential proportion for CHIS 2001	3-8
4-1	Number of households with a completed screener interview	4-2
4-2	Variables considered for the creation of nonresponse adjustment cells for the adult weights	4-3
4-3	Number of adult completed screener interviews with interruptions in telephone service	4-7
4-4	Common variables between CHIS 2001 and the March 2001 CPS	4-8
4-5	Nontelephone adjustment cell definition for CHIS 2001	4-10
6-1	Number of completed adolescent and adolescent insurance interviews	6-1
6-2	Number of trimmed weights and range of trimming factor for the adolescent and adolescent insurance weights	6-3
7-1	Definitions of the dimensions used in raking	7-3
7-2	Overall adjustment raking factors for adult, child, adolescent and adolescent insurance interviews	7-6
7-3	Population in California	7-8
7-4	SF1 available group quarter counts	7-9
7-5	Computing the control totals for race.	7-10

TABLE OF CONTENTS (continued)

List of Tables (continued)

Table		Page
7-6	Number of people in group quarters by race alone or in combination with one or more races	7-12
7-7	Dimension 11 control totals	7-13
8-1	Description of imputed variables	8-2
8-2	Number and percentage of completed adult interviews with missing county	8-4
8-3	Distribution of self-reported strata versus sampling strata	8-6
8-4	Number and percentage of completed interviews with missing self-reported sex and age	8-7
8-5	Number and percentage of completed interviews with missing self-reported race and/or ethnicity	8-9
8-6	Counts and percentages of imputed self reported race and ethnicity	8-10
9-1	CHIS 2001 race-ethnic supplemental samples	9-4
9-2	Race-ethnic group of sampled persons by type of interview	9-9
9-3	Size of the race-ethnic adjustment factor	9-11
9-4	Dimensions used in raking	9-14
9-5	Effective sample sizes for CHIS 2001 supplemental samples- adult interview	9-21
10-1	Average DEFF and DEFT for estimates from the adult interview	10-5
10-2	Average DEFF and DEFT for estimates from the child interview	10-6
10-3	Average DEFF and DEFT for estimates from the adolescent interview	10-8

TABLE OF CONTENTS (continued)

List of Appendixes

Appendix A

A-1	Household weighting for the combined RDD sample and geographic supplemental samples by stratum	A-1
A-2	Extended interview weighting for adult interview by stratum (combined RDD sample and geographic supplemental samples)	A-11
A-3	Extended interview weighting for child interview by stratum (combined RDD sample and geographic supplemental samples)	A-16
A-4	Extended interview weighting for adolescent interview by stratum (combined RDD sample and geographic supplemental samples)	A-21
A-5	Extended interview weighting for adolescent insurance interview by stratum (combined RDD sample and geographic supplemental samples)	A-26

Appendix B

B-1	Average DEFF and DEFT for Estimates from the Adult Interview for Continuous and Categorical Variables	B-1
B-2	Average DEFF and DEFT for Estimates from the Child Interview for Continuous and Categorical Variables	B-3
B-3	Average DEFF and DEFT for Estimates from the Adolescent Interview for Continuous and Categorical Variables	B-5

Appendix C

C-1	Race-ethnic supplemental samples	C-1
C-2	Sample size, sums of weights and coefficient of variation for the race-ethnic samples	C-7
C-3	Variables used in the evaluation of the weights	C-8

1. CHIS 2001 DESIGN AND METHODOLOGY SUMMARY

1.1 Overview

The 2001 California Health Interview Survey (CHIS 2001) is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. The focus of the survey is on a variety of public health topics, including access to health care and health insurance coverage. CHIS 2001 is the largest state health survey ever undertaken in the United States. It is a random digit dialing (RDD) telephone survey of California households designed to produce reliable estimates for the whole state, for large- and medium-sized population counties in the state, and for groups of the smallest population counties. Three California cities that have their own health departments were also sampled as part of CHIS 2001.

The survey design supports study of California's major race and ethnic groups, and a number of smaller ethnic groups within the state. Adults, parents of children below age 12, and adolescents (ages 12-17) residing in California households are the eligible respondents to the survey. CHIS 2001 collected data between November 2000 and October 2001. The plans are to conduct independent cross-sectional surveys of the California population on a biannual basis to monitor important health-related indicators and potentially track changes over time. CHIS 2001 is the first of these planned surveys.

CHIS 2001 collected information on if, where, and how people get health care in California. The goal is to provide health planners, policymakers, state, county, and city health agencies, and community organizations with information on the health and health care needs facing California's diverse population. For example, the number and characteristics of adults, children, and adolescents without access to care and lacking health insurance can be estimated from the data collected in CHIS 2001. Other key estimates on the prevalence of cancer screening, diabetes, asthma, and other health conditions can also be produced. The survey includes major content areas, such as health status and conditions, health-related behaviors, access to health care services, and health insurance coverage.

1.2 Sample Design Objectives

The CHIS 2001 sample is designed to meet two objectives: (1) provide local-level estimates for counties and groupings of counties with populations of 100,000 or more; and (2) provide statewide estimates for California's overall population and its larger race/ethnic groups, as well as for several smaller ethnic groups. To address these objectives, the sample was allocated by county and aggregates of smaller counties, with supplemental samples of selected populations and cities. Table 1-1 shows the 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.

Table 1-1. California county and county group strata used in the 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, 2001 California Health Interview Survey.

Samples were also drawn from each of the three California cities that have their own local health department. In addition, supplemental samples were developed for three counties that contracted for additional sample to enhance their overall estimates. These city and supplemental county samples were in the following locations:

- The cities of Berkeley, Long Beach, and Pasadena; and
- The counties of San Francisco, Santa Barbara, and Solano.

The three city samples and the Solano county supplemental sample were implemented with and incorporated in the original statewide RDD sample. The separate San Francisco and Santa Barbara

supplemental samples were subsequently added to the statewide RDD sample prior to constructing the sample weights and are part of the final CHIS 2001 RDD sample file.

To accomplish the second objective, larger sample sizes were allocated to the more urban counties where a significant portion of the state's African American and Asian ethnic populations reside. Additionally, supplemental samples were used to improve the sample size and precision of the estimates for specific ethnic groups. The supplemental ethnic group samples in CHIS 2001 were as follows:

- South Asian, Cambodian, Japanese, Korean, and Vietnamese;
- American Indian/Alaska Natives in urban and rural areas; and
- Latinos residing in Shasta County (a sample requested by the local health department).

1.3 Data Collection

To capture the rich diversity of the California population, interviews were conducted in six languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, Korean, and Khmer (Cambodian). These languages were chosen based on research that identified these as the languages that would cover the largest number of Californians in the CHIS sample design 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 2001 data collection for the CHIS project. Westat staff interviewed one randomly selected adult in each sampled household. In those households with children (under age 12) or adolescents (ages 12-17), 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 who knew the most about the health and care of the sampled child was interviewed. The sampled adolescents responded for themselves, but only after a parent or guardian gave permission for the interview. Since adolescents were not reliable sources concerning their own health insurance coverage, the parents of sampled adolescents were interviewed about this topic separately.

One criterion for the adolescent and child to be selected for the survey is that they had to be "associated" with the selected adult. This meant that in most cases the interviewed adult had to be either the parent or guardian. The CHIS 2001 sample weights adjust for this selection criterion so as not to bias

estimates based on the adolescent and child surveys. Table 1-2 shows the number of completed adult, child, adolescent, and adolescents' health insurance interviews in CHIS 2001, by the type of sample (RDD or supplemental sample).

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

Type of sample	Adult	Child	Adolescent	Adolescent insurance
Total RDD + supplemental cases	57,848	13,276	6,058	8,302
RDD (includes 3 cities + Solano county supplemental cases)	54,122	12,392	5,733	7,809
Santa Barbara supplemental cases	206	49	22	31
San Francisco supplemental cases	1,100	151	46	79
Total CHIS 2001 RDD file	*55,428*	*12,592*	*5,801*	*7,919*
Other supplemental samples:				
South Asian	443	158	39	65
Cambodian	126	44	37	44
Japanese	330	51	18	33
Korean	326	95	30	44
Vietnamese	540	124	34	60
American Indian/Alaska Native	351	106	51	71
Shasta Latinos	304	106	48	66

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

The interviews done in English were administered using Westat's computer-assisted telephone interviewing (CATI) system. Spanish and Vietnamese language interviews were also conducted entirely in CATI, while interviews conducted in Cantonese, Mandarin, Korean, and Khmer used English CATI screens and paper translations in tandem. The average adult interview took around 32 minutes to complete. The average child and adolescent interviews took 14 minutes and 19 minutes, respectively. Interviews in the non-English languages generally averaged longer to complete. Approximately 12 percent of the adult interviews were completed in a language other than English, as were 21 percent of all child (parent proxy) interviews and 9 percent of all adolescent interviews.

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

Table 1-3. Survey topic areas by instrument

Adult interview	Child interview	Adolescent interview
Age, sex, race, ethnicity	Age, sex, race, ethnicity	Age, sex, race, ethnicity
Physical activity		Physical activity
	Bike helmet use	Bike helmet, seatbelt use
	Recent serious injury	Recent serious injury
Health status	Health status	Health status
Women's health	Child care	
Chronic health conditions	Asthma, ADD	Asthma, diabetes
Cancer history, screening		
Skin cancer prevention	Skin cancer prevention	Skin cancer prevention
Health care use and access	Health care use and access	Health care use and access
Alcohol, tobacco use		Alcohol, tobacco, drug use
Mental health		Mental health
Health insurance	Health insurance	Health insurance
Diet (fruit-vegetable intake)	General diet	General diet
Dental health	Dental health	Dental health
Employment		Employment
Gun access, training		Gun access, violence
Income		
	Family interaction	Parental involvement
	Video games, computer use	Video games, computer use
Sexual orientation		Sexual behavior, orientation
		Future plans

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

1.4 Response Rate

The overall response rate for CHIS 2001 is a composite of the screener completion rate (i.e., success in introducing the survey to a household in order to select a respondent), and the extended interview completion rate (i.e., success in getting the selected respondent to complete the full interview). For the adult survey, the screener completion rate was 59.2 percent and the extended interview completion rate was 63.7 percent. This gives an overall response rate of 37.7 percent. To maximize the survey's response rate, an advance letter (in five languages) was mailed to all sampled telephone numbers for which an address could be obtained from reverse directory services. Approximately 66 percent of the sample was mailed an advance letter. Response rates varied by sampling stratum and were slightly higher in households that received an advance letter.

To assist in achieving sample size goals, respondents that completed 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." This resulted in 397 "partial completes" being included in the final adult survey data. Employment and income information as well as potential public program eligibility and food insecurity information would be missing from these cases.

Proxy interviews were allowed for frail and ill persons over the age of 65. The reason is that health estimates made for elderly persons could be biased if this is not allowed. Eligible selected persons were recontacted and offered a proxy option and 316 had a proxy interview 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 Random Digit Dial Sample

To produce correct 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. Sample weighting was carried out in CHIS 2001 to accomplish the following objectives:

- Compensate for differential probabilities of selection for households and persons (Note: households with listed addresses and thus eligible for an advance letter were assigned a probability of selection of 1.25 over unlisted households);
- 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 for the RDD samples (each stratum is an independent sample), a household weight was created for all households 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 each of the following:

- Subsampling for listed address/advance letter status;
- Unknown residential status;
- Screener interview nonresponse;
- Multiple telephone numbers; and
- Household poststratification.

A "poststratified household weight" was then used to compute a person-level weight. This person-level weight incorporates the within-household probability of selection of the sampled person and adjusts for nonresponse, plus an adjustment resulting from raking the data to person-level control totals. Each of these adjustments corresponds to a multiplicative weighting factor.

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 were adjusted, the process was iterated until the control totals for all the dimensions were simultaneously satisfied (within a specified tolerance).

There are 11 dimensions used in CHIS 2001. The first 10 dimensions are created by combining demographic variables (age, sex, race, and ethnicity) and different geographic areas (city, county, group of counties, and state). The 11th dimension is created to adjust the weights for households without a telephone number.

The control totals used in the raking were derived from the *Census 2000 Summary File 1* (SF1). Population items in SF1 include sex, age, race, ethnicity (Latino/non-Latino), household relationships, and group quarters. The race classification in SF1 include six groups: White, African American, American Indian/Alaska Native, Asian, Native Hawaiian/Pacific Islander, and a category of Other Race. Since a person could report multiple races, the SF1 provided counts for each of 63 possible race combinations a person could report.

One of the limitations of using the SF1 for the control totals is the inability to produce counts that exclude the fraction of the population living in "group quarters" (e.g., nursing homes, prisons) for

some dimensions used in CHIS 2001. The group quarter population represented 2.4 percent of the total population in California. As a result, the number of persons living in group quarters was estimated for some of the raking dimensions, and the SF1 totals were reduced by these estimated amounts prior to raking.

1.6 Imputation Methods

Three different imputation procedures were used in CHIS 2001 to fill in missing responses that were essential for weighting the data or for such basic descriptive purposes as income categories. The first imputation technique is deterministic or non-stochastic in nature. Deterministic imputation was used to fill in the missing items for self-reported county of residence (item AH42). These imputations required no randomization because other geographic data are available that can be used to determine the respondent's county of residence with a relatively high level of probability of being correct although not with 100 percent certainty in all cases.

The second 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.

The third technique is hotdeck imputation. Hotdeck imputation was used to impute race, ethnicity, and household income in CHIS 2001. The hotdeck approach is probably the most commonly used method for assigning values for missing responses in large-scale household surveys.

With a hotdeck, 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. To carry out hotdeck imputation for CHIS 2001, 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 household structure. The recipient is then randomly imputed the same household income, ethnicity/race (depending on the items that need to be imputed) from one of the donors in the pool. Once a donor is used, it is removed from the pool of donors.

Imputation flags are used in the data file 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 2001:

- Report 1 - Sample Design
- Report 2 - Data Collection Methods
- Report 3 - Data Processing Procedures
- Report 4 - Response Rates
- 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 FOR RANDOM-DIGIT DIALING AND GEOGRAPHIC SAMPLE WEIGHTS

This chapter introduces the concept of weighting and provides some background on the weights developed for analyzing CHIS 2001 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 and the base weight 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 RDD sample and supplemental samples, the chapter begins with the general reasons why the fully adjusted weights should be used. It also describes the details, advantages, and disadvantages of weighting. A list of the weights created for CHIS 2001 is also included. In CHIS the RDD sample and geographic supplemental samples are combined and weighted together. In contrast, each of the race-ethnic supplemental samples is weighted separately, for reasons described later in Chapter 9.

2.1 Weighting Approach

In order to produce estimates, weights are applied to 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 for households and persons;
- Reduce biases occurring because nonrespondents may have different characteristics than respondents;
- Adjust, to the extent possible, for undercoverage in the sampling frames and in the conduct of the survey; and
- Reduce the variance of the estimates by using auxiliary information.

In CHIS 2001 one set of weights was created for each instrument or questionnaire (adult, child, adolescent, and adolescent insurance) for the combined RDD and geographic supplemental samples. For each of the seven race-ethnic supplemental samples another set of weights was created for each instrument. The sets of weights for each type of sample are shown in Table 2-1. The levels of estimates (i.e., state or county level) that can be produced using these weights are also indicated in Table 2-1.

Table 2-1. CHIS 2001 weights and type of estimate

Type of sample	Adult	Child	Adolescent	Adolescent insurance
RDD*	C, S	C, S	C, S	C, S
American Indian/ Alaska Native	S, U	S, U	S, U	S U
Cambodian	S	S	S	S
South Asian	S	S	S	S
Japanese	S	S	S	S
Korean	S	S	S	S
Vietnamese	S	S	S	S
Shasta Latinos	C	C	C	C

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Where C: Estimates at the county/stratum-level

S: Estimates at the state-level

U: Estimates for urban/rural areas

* Includes geographic supplemental samples

Each final weight is the result of a series of sequential adjustments made using the base weights. However, this process for the combined RDD and geographic samples differed from that used for the race-ethnic samples. The details for the creation of the weights for the combined RDD-geographic sample are given in Chapters 3, 4, 5, and 6, while Chapter 9 presents the information for race-ethnic supplemental samples.

As part of the weighting process for the RDD samples, a household weight is created for all households 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 listed/mail status;
- Unknown residential status;

- 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 adjustments incorporate the within-household probability of selection of the sampled person and account for nonresponse. Each of the adjustments 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 Report 1: Sample Design). For these weights, the factors include:

- Section H adult extended interview nonresponse;
- Probability of selection of the adult and his or her spouse;
- Probability of selection of the child or adolescent;
- Extended child, adolescent, and adolescent insurance 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. A description of the imputation process and the imputed variables is described in Chapter 8. Chapter 9 describes the creation of the weights for the race-ethnic supplemental samples. Intermediate weights are created as the product of the base weight and all the factors that reflect any subsampling either at the household or person level. The final weight is then computed by raking the intermediate weight to control totals for the race or ethnic group. Chapter 10 describes the methods for variance estimation for CHIS 2001.

Appendix A contains tables that show the effect of virtually each step of the weighting process at the household and person levels. Throughout this report we refer to specific tables and rows in Appendix A 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 the estimates from the survey will be biased. In CHIS 2001, 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 2001. In this procedure, nonresponse adjustments are computed and applied separately by cell, where a cell is defined using characteristics known for both nonrespondents and respondents. For example, we know from the telephone number the county in California that it is associated with, 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 adjustment reduces bias if 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 2001 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 the defined 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 is 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 because uncovered units or persons were never eligible to be sampled. The procedure used to adjust for undercoverage is to use 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 2001 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 2001, the control totals are derived from the Census 2000 Summary File 1 for California published by the U.S. Census Bureau. Totals for the number of households with or without someone under 18 years old by sampling strata 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 2001 combined RDD and geographic supplemental samples. The weighting for the race-ethnic supplemental samples is presented in Chapter 9.

3. HOUSEHOLD WEIGHTING

The first step in the weighting process for the combined RDD and geographic supplemental samples of CHIS 2001 is creating a household weight for each completed screener interview. Although the household weight is not used for analytical purposes because the screener captured data mainly for sampling purposes, this weight is a key element for the computation of the person weights (i.e., adult, child, adolescent, and adolescent insurance).

This chapter is divided into six 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 weights. These adjustments account for subsampling based on listed/mail status, 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 2001 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 any details in this report.

3.1 Base Weights

The RDD sample was drawn using a list-assisted approach from a stratified frame of 100 banks[1] with at least one listed telephone number in the state of California. Using this approach, a bank is drawn for the frame and two digits are randomly generated to complete the sampled telephone number. The base weight of a telephone number is then computed as the inverse of the probability of selecting the number, that is the ratio of the total number of 100 banks in the strata multiplied by 100 to the number of telephone numbers sampled.

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

This weight is appropriate when only one sample is drawn or when there are no changes in the number of banks in the frame. The CHIS 2001 RDD sample was drawn at two different points in time (July 2000 and February 2001). During this period there were changes in the frame. New exchanges and banks appeared in the frame while others were removed. Since the number of banks was not the same when the samples were drawn we used the average number of banks to compute the base weight. The base weight $HHBSW_i$ for the *i*-th telephone number in a given stratum is[2]

$$HHBSW_i = \frac{\bar{N} \cdot 100}{n_1 + n_2}$$

where

n_1 = the number of telephone numbers sampled in July 2000;
n_2 = the number of telephone numbers sampled for February 2001; and
$\bar{N}$ = the average number of banks in July 2000 and February 2001. This quantity is

$$\bar{N} = \frac{N_1 + N_2}{2}$$

where N_1 is the total number of banks in the frame in July 2000 for the stratum and N_2 is the total number of banks in the frame in February 2001 for the stratum.

The expression above applies to all the strata except for the cities (Pasadena, Long Beach, remainder of Los Angeles, Berkeley, and remainder of Alameda) and the supplemental samples for San Francisco and Santa Barbara. In the case of cities, the initialsample selection (July 2000) was drawn from the combined strata (i.e., Los Angeles County or Alameda County). In the second selection, the sample was drawn from the exchanges in separate strata, each representing the city or the remainder of the county. In this case, the base weight $HHBSW_i$ is computed as

$$HHBSW_i = \frac{N_2 \cdot 100}{n_1^* + n_2}$$

[2] As a notational convenience, a subscript for the sampling stratum is omitted in the description of the weights.

where n_1^* is the realized January 2001 number of sampled telephone numbers in the newly-defined stratum for the city.

The San Francisco and Santa Barbara supplemental samples were drawn at a later time (August 2001). These two supplemental samples came from a different subset of banks from the ones already sampled for these areas in the RDD sample. The expression of the base weights for San Francisco and Santa Barbara is similar to the base weight used for the race-ethnic supplemental samples described in Chapter 9. We delay the discussion of the development of the base weight for these samples until that chapter.

Table 3-1 gives the number of banks, sample cases, and base weights by stratum. Table 3-2 shows the supplemental sample cases for the San Francisco and Santa Barbara samples for CHIS 2001. Table A-1 in Appendix A (rows 1.1 and 1.2) lists the sample counts and sums of base weights by sampling stratum.

3.2 Subsampling Adjustments

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 prior to the start of the data collection period. 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, a telephone number was classified by whether a mailing address could be associated with the telephone number[3] (i.e., mail status). We refer to those telephone numbers that were associated with a mailing address as having a "mailable" address. A second procedure involved a tritone purging method to identify the working status of a telephone number (working or nonworking). Telephones classified as nonresidential or nonworking were not dialed for CHIS 2001.

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

Table 3-1. Number of banks for CHIS 2001 sample selection, average number of banks, sample size, and base weight by sampling strata

Stratum number	Description	July 2000 banks (N_1)	February 2001 banks (N_2)	Average number of banks ($\bar{N}$)	July 2000 sample size (n_1)	February 2001 sample size (n_2)	Base weight
1.1	Long Beach	74,350	3,520	3,520.0	1,431[1]	4,080	63.87
1.2	Pasadena		1,961	1,961.0	823[1]	5,281	32.13
1.3	Remainder Los Angeles		70,399	70,399.0	29,096	45,247	94.69
2	San Diego	19,937	20,275	20,106.0	7,350	8,276	128.67
3	Orange	23,913	24,091	24,002.0	7,200	10,861	132.89
4	Santa Clara	15,578	15,943	15,760.5	4,200	6,679	144.87
5	San Bernardino	10,800	10,967	10,883.5	4,200	4,407	126.45
6	Riverside	10,167	10,504	10,335.5	3,750	3,957	134.11
7.1	Berkeley	13,234	1,157	1,157.0	298[1]	4,719	23.06
7.2	Remainder of Alameda		12,436	12,436.0	3,152	4,663	159.13
8	Sacramento	10,093	10,446	10,269.5	3,450	3,634	144.97
9	Contra Costa	8,308	8,682	8,495.0	3,300	4,054	115.52
10	Fresno	6,293	6,340	6,316.5	2,850	4,341	87.84
11	San Francisco	10,070	10,173	10,121.5	2,700	5,388	59.25[2]
12	Ventura	5,156	5,485	5,320.5	2,850	2,975	91.34
13	San Mateo	6,655	6,679	6,667.0	3,150	3,788	96.09
14	Kern	5,070	5,112	5,091.0	2,850	3,171	84.55
15	San Joaquin	3,695	3,739	3,717.0	2,850	2,559	68.72
16	Sonoma	4,099	4,245	4,172.0	1,800	2,165	105.22
17	Stanislaus	2,975	2,983	2,979.0	1,800	2,141	75.59
18	Santa Barbara	3,202	3,358	3,280.0	1,800	2,516	62.97[2]
19	Solano[3]	2,686	2,751	2,718.5	1,800	6,161	34.15
20	Tulare	2,689	2,701	2,695.0	1,800	3,222	53.66
21	Santa Cruz	2,510	2,525	2,517.5	1,800	3,006	52.38
22	Marin	3,010	3,132	3,071.0	1,800	3,181	61.65
23	San Luis Obispo	2,189	2,232	2,210.5	1,800	2,335	53.46
24	Placer	2,159	2,298	2,228.5	1,800	2,480	52.07
25	Merced	1,137	1,159	1,148.0	1,650	2,444	28.04
26	Butte	1,495	1,532	1,513.5	1,800	1,678	43.52
27	Shasta	1,363	1,358	1,360.5	1,800	1,966	36.13
28	Yolo	1,192	1,177	1,184.5	1,800	1,954	31.55
29	El Dorado	1,208	1,233	1,220.5	1,650	2,775	27.58
30	Imperial	711	721	716.0	1,800	2,147	18.14
31	Napa	1,094	1,096	1,095.0	1,800	2,757	24.03
32	Kings	664	665	664.5	1,800	2,506	15.43
33	Madera	743	780	761.5	1,800	2,291	18.61
34	Monterey, San Benito	3,738	3,791	3,764.5	1,650	3,783	69.29
35	Del Norte, Humboldt	1,453	1,460	1,456.5	1,800	3,234	28.93
36	Lassen, Modoc, Siskiyou, Trinity	976	983	979.5	1,800	3,456	18.64

Table 3-1. Number of banks for CHIS 2001 sample selection, average number of banks, sample size, and base weight by sampling strata (continued)

Stratum number	Description	July 2000 banks (N_1)	February 2001 banks (N_2)	Average number of banks ($\bar{N}$)	July 2000 sample size (n_1)	February 2001 sample size (n_2)	Base weight
37	Lake, Mendocino	1,201	1,209	1,205.0	1,800	2,923	25.51[4]
38	Colusa, Glen, Tehama	732	733	732.5	1,650	2,457	17.84[4]
39	Sutter, Yuba	1,013	1,022	1,017.5	1,800	2,557	23.35
40	Plumas, Nevada, Sierra	1,148	1,174	1,161.0	1,800	2,735	25.60
41	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	2,013	2,038	2,025.5	1,800	3,352	39.31
	Total	270,719	276,265	274,436.5	131,700	200,302	

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

[1] Realized January 2001 number of sampled telephone numbers in the city strata (n_1^*).

[2] Average base weight. See Chapter 9 for the description of how these base weights are computed.

[3] Includes geographic supplemental samples in this area

[4] Base weights of 20.47 and 23.63 were accidentally used for strata 37 and 38 respectively, due to a numerical transposition. They were rectified at this stage with the correct weights of 25.51 and 17.84 by the poststratification stage because it only involved multiplication by a constant within stratum. See Table A-1 in the appendix.

Table 3-2. Sample sizes for San Francisco County and Santa Barbara County supplemental samples

Supplemental sample	Number of banks	Sample size
San Francisco County	11,014	12,241
Santa Barbara County	1,031	896
Total	12,045	13,134

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

The efficiency of the sample was further improved by stratifying the telephones numbers and subsampling based on the strata at different rates (Brick, Judkins, Montaquila, and Morganstein, 2002). Listed and/or mailable telephone numbers are more likely to be residential so the cost of finding a residence is much lower in the substratum of listed/mailable numbers. In addition, households with listed/mailable telephone numbers are more likely to cooperate with most surveys.[4] 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.

[4] 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.)

In CHIS 2001 substrata were created using the original strata and the information on working status (residential, business, or nonworking telephone number), the listed status (telephone number listed or not), and mail status (the telephone number has a "mailable" address or not). As described in Report 1: Sample Design, the subsampling substrata used were different in the July 2000 and February 2001 samples as shown in Table 3-3. The telephone numbers not selected in the subsample were eliminated and never dialed (the only reason these were subsampled was because of the sequence of steps used). Table A-1 (rows 2.5a through 2.5f) shows the observed sampling rates for the July 2000 and February 2001 samples.

Table 3-3. Substratum definition for subsampling and planned subsampling rate

RDD sampling frame	Sub-strata	Description	Planned subsampling rate
July 2000	L	Listed as residential	1.0
	NL	Unlisted	0.8
	NR1	Nonresidential (business and nonworking)	0.8
February 2001	M	Mailable address	1.0
	NM	No mailable address	0.8
	NR2	Nonresidential (business and nonworking)	0.8

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

The subsampling adjusted household weight, $HHA1W_i$, that accounts for this subsampling is

$$HHA1W_i = HHA1F_c \cdot HHBSW_i ,$$

where

$$HHA1F_c = \begin{cases} \dfrac{\sum_{i \in REVSTR=c} HHBSW_i}{\sum_{i \in INSMP \, and \, i \in REVSTR=c} HHBSW_i} & \text{if } i \in INSMP \text{ and } i \in REVSTR = c \\ 0 & \text{otherwise} \end{cases} .$$

In this notation, *INSMP* is the set of telephone numbers retained in the subsample, and c refers to the substrata cells defined in Table 3-3. The subsampling weighting adjustment is done within sampling strata. Table A-1 (row 2.3) shows the sum of the weights after the subsampling adjustment. Two points are worth noting. First, the telephones in substrata NR1 and NR2 (those identified in the tritone

purge as nonresidential) were dropped from the weighting process at this point because they are not residential numbers. Second, the San Francisco and Santa Barbara supplemental samples were not subsampled by listed/mailable status.

3.3 Unknown Residential Status Adjustment

At the end of data collection not all telephone numbers can be classified as residential despite being dialed many times. These numbers are considered as unknown residential status. They are telephone numbers that reached only answer machines (screener disposition code of NM) or were never answered (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 2001, the estimated proportion of unknown residential telephone numbers considered residential (p_{res}) 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. The proportion (p_{res}) was computed within groups defined by urban status,[5] mail status of the telephone number, and the answering machine status given by the interviewer. Table 3-4 shows the values of p_{res} computed using the survival analysis method.

As expected, the estimated proportion of residential households decreases when an answering machine is 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 91.8 percent, while the estimated proportion of those coded as urban-mailable-nonresidential is 23.4 percent. The lowest percentages of residential telephone numbers are for the numbers that were not mailable and were never answered or had answering machine messages coded as nonresidential.

[5] For this purpose, urban strata are defined as those strata with a population of over 500,000 persons. Stratum 1.1 (Long Beach) through Stratum 15 (San Joaquin) in Table 3-1 are urban, while the remaining strata are rural.

Table 3-4. Estimated residential proportion for CHIS 2001

Urban status	Mail status	Answering machine code	p_{res}
Urban	mailable	no machine	0.49640
Urban	mailable	residential	0.91810
Urban	mailable	nonresidential	0.23375
Urban	mailable	unknown	0.81239
Urban	non-mailable	no machine	0.08244
Urban	non-mailable	residential	0.79869
Urban	non-mailable	nonresidential	0.06335
Urban	non-mailable	unknown	0.42212
Rural	mailable	no machine	0.61621
Rural	mailable	residential	0.94073
Rural	mailable	nonresidential	0.32248
Rural	mailable	unknown	0.82476
Rural	non-mailable	no machine	0.24183
Rural	non-mailable	residential	0.87512
Rural	non-mailable	nonresidential	0.05862
Rural	non-mailable	unknown	0.54502

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

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, $HHA2W_i$, for the RDD sample is

$$HHA2W_i = HHA2F_i \cdot HHA1W_i$$

where

$$HHA2F_i = \begin{cases} \dfrac{\sum_{i \in RES} HHA1W_i + \sum_{i \in UNK\_RES} p_{res} \cdot HHA1W_i}{\sum_{i \in RES} HHA1W_i} & \text{if } i \in RES \\ 0 & \text{if } i \in UNK\_RES \end{cases}$$

where the subscript *RES* denotes telephone numbers identified as residential and *UNK_RES* denotes telephone with unknown residential status.

The adjustment factor given above for the RDD sample does not apply to the San Francisco and Santa Barbara supplemental samples because the sampling methods are different. For the San Francisco and Santa Barbara samples, the adjustment factor, $HHA2F_i$, is

$$HHA2F_i = \begin{cases} \dfrac{\sum_{i \in RES} HHA1W_i + \sum_{i \in IN} HHA1W_i + \sum_{i \in UNK\_RES} p_{res} \cdot HHA1W_i}{\sum_{i \in RES} HHA1W_i + \sum_{i \in IN} HHA1W_i} & \text{If } i \in RES,\ IN \\ 0 & \text{If } i \in UNK\_RES \end{cases}$$

where *IN* denotes telephone numbers that were ineligible because they were not in San Francisco or Santa Barbara in these supplemental samples.

In the supplemental samples, residential telephone numbers not located in the San Francisco or Santa Barbara were coded as ineligible [6] and no interview was conducted. When adjusting for the portion of households that are residential among those with unknown residential status, we assume that some of these households are not located in San Francisco or Santa Barbara. Through this adjustment, the weights of the telephone numbers coded as unknown residential were distributed among the eligibles and ineligibles (eligibility based on geography) in the same proportion as observed in the sample. After this adjustment, only eligible residential households were retained for the next weighting adjustment. Although the sum of weights of the ineligibles is not zero, they were removed from the weighting process and were not further adjusted after this step because no interviews were conducted with these households.

While the value of p_{res} is computed for the groups defined in Table 3-4, the weight adjustment is made in cells within sampling stratum and type of sample (RDD, San Francisco, and Santa Barbara samples). The cells were created using the mail and listed status of the telephone number. Table A-1 (rows 3.2 and 3.3) gives the sum of weights before and after making the adjustment for unknown residential status.

3.4 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

[6] Unlike the screener questionnaire used for the RDD, the screener questionnaire for San Francisco and Santa Barbara included an eligibility question (SC4) based on geography. The respondent was asked if he or she was located within these areas. If the answer was yes, the interview continued. Otherwise, the interviewer ended the interview and coded the case as ineligible (SCRNRSLT=IS).

not complete the screener interview. For the RDD sample, the screener nonresponse adjusted household weight, $HHA3W_i$, is

$$HHA3W_i = HHA3F_c \cdot HHA2W_i ,$$

where

$$HHA3F_c = \begin{cases} \dfrac{\sum_{i \in SC\_R, SC\_NR} HHA2W_i}{\sum_{i \in SC\_R} HHA2W_i} & i \in SC\_R \\ 0 & i \in SC\_NR \end{cases} ,$$

and *SC_R* is the set of screener respondents, *SC_NR* is the set of screener nonrespondents, and c is the indicator for the screener nonresponse adjustment cell.

The nonresponse adjustment cells for this adjustment were created using the telephone mail status within sampling stratum. These cells have different response rates due to the effect of the pre-notification letter sent to households with a known address and due to differences associated with the mailable and nonmailable groups. Table A-1 (rows 4.1 and 4.2) gives the sum of weights before and after the screener nonresponse adjustment.

3.5 Multiple Telephone Adjustment

At the end of the screener interview, the interviewer collected information about the existence of additional telephone numbers and their use in the household (screener interview questions SC9 and SC10). If the additional telephone number was used for residential purposes (telephone not used solely for business, computer use, etc.), then the household had a greater probability of selection because it could have been selected through the other number. For these households, the household weight is adjusted to reflect the increased probability of selection. The multiple telephone adjusted household weight, $HHA4W_i$, is

$$HHA4W_i = HHA4F_i \cdot HHA3W_i ,$$

where

$$HHA4F_i = \begin{cases} 0.5 & \text{if household } i \text{ has more than one residential telephone number} \\ 1 & \text{otherwise} \end{cases}$$

In this adjustment we assume that there is at most one additional telephone number. 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), while the weight of other the telephone number (with the completed interview) was adjusted to account for the two numbers. Table A-1 (rows 5.1 and 5.2) identifies the sum of weights before and after the multiple telephone adjustment.

3.6 Household Poststratification

The final step in weighting the screener interviews was to poststratify the household weights to household control totals from the Census 2000 data (Census 2000 Summary File 1 for California released by the U.S. Census Bureau). The poststratification cells were created for households with and without individuals under 18 years old by sampling stratum. The poststratification adjustment reduces potential bias related to different response rates for households with or without children and/or adolescents.

The household poststratification weight, $HHA5W_i$, is

$$HHA5W_i = HHA5F_k * HHA4W_i$$

where

$$HHA5F_k = \frac{CNT_k}{\sum_{i \in k} HHA4W_i}$$

where CNT_k is the control total for cell k.

The overall poststratification adjustment factors for the state and by sampling stratum are listed in Table A-1 (row 6.4). The magnitude of this adjustment is sometimes used as a measure of the undercoverage of the estimate of the total number of households. In this case, there are other factors included in the adjustment so that it is not a very precise measure of coverage. For example, the sampling strata do not always correspond to the counties the respondents live in because the telephone numbers are not completely nested in the geographic areas.

After examining the poststratification factors, some combining of cells was determined to be appropriate to reduce the size of the factors. The cells for the sampling strata within Los Angeles County (Long Beach, Pasadena, and remainder of Los Angeles County) were combined, and the sampling strata from Alameda County (Berkeley and remainder of Alameda County) were combined because of the size of the factors. This combination is also consistent with the fact that the differences between the assignment of the sampling strata and the actual location of the interviewed household are most pronounced in these areas. As a result of the combining, the sum of weights for these strata does not add up to control totals from the 2000 Census when considered separately, but does when combined at the county-level. The sum of weights before and after household poststratification are in Table A-1 (rows 6.2 and 6.3).

4. ADULT WEIGHTING

An adult final weight was created for all adults that completed or partially completed the adult extended interview.[1] 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 that could not be interviewed because they reside in nontelephone households, the raking adjustment was modified to reduce the bias from this source. Details on creating the adult weights follow.

4.1 Adult Initial Weight

As described in Report 1: Sample Design, adults were sampled with different probabilities of selection. For most households, the adult was sampled with equal probability from all listed adults. Differential probabilities of selecting the adults were used only in households where there were 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 (see Report 1: Sample Design).

The initial adult weight is the product of the household weight and the inverse of the probability of selection of the adult. The expression for the adult initial weight, $ADA0W_i$, is

$$ADA0W_i = \frac{1}{ADPROB_i} \cdot HHA5W_i,$$

[1] Adult extended interviews that are considered complete have disposition codes of "CA" or "CP." CP includes all the partially completed adult interviews, i.e., interviews that were completed through Section I on health insurance of the extended adult interview.

where

$$ADPROB_i = \frac{ADMOS_i}{\sum_i ADMOS_i},$$

and $ADPROB_i$ is the probability of selecting the sampled adult within the household, and $ADMOS_i$ is the measure of size for adult *i*. The measure of size for adult *i* is

$$ADMOS_i = \begin{cases} 2 & \text{if the adult } i \text{ is sampled at the lower rate} \\ 1 & \text{otherwise} \end{cases}.$$

Table 4-1 shows the distribution of households where the screener interview was completed and classified by the way the sample adult was selected (equal or differential probability of selection). In almost 90 percent of the households where the screener interview was completed, the adult was sampled with equal probability.

Table 4-1. Number of households with a completed screener interview*

Sample adult	Number of households with completed screener interview	Percentage
Equal probability	74,318	88.4
Differential probability	9,733	11.6
Total	84,051	100.0

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.

4.2 Adult Nonresponse Adjustment

In some households the screener interview was completed but the sampled adult did not complete the extended adult interview. To account for sampled adults that did not complete the extended interview, we adjusted the adult initial weight for extended interview nonresponse. The adult nonresponse adjusted weight, $ADA1W_i$, is

$$ADA1W_i = ADA1F_c \cdot ADA0W_i .$$

where

$$ADA1F_c = \begin{cases} \dfrac{\sum_{i \in ER,IN,NR} ADA0W_i}{\sum_{i \in ER,IN} ADA0W_i} & \text{if } i \in ER, IN \\ 0 & \text{otherwise} \end{cases}$$

and *ER* is the set of eligible respondents, *IN* is the set of ineligibles, *NR* is the set of nonrespondent; and *c* indicates the adult nonresponse adjustment cell.

The variables listed in Table 4-2 are the ones that were considered as candidates to create the nonresponse adjustment cells. Analyses showed the response rates by sex and age groups were the most variable and best candidates for the cells. As a result, the nonresponse cells were created by classifying adults into sex and age groups. Cells with less than 30 respondents or with large adjustment factors were combined with adjacent groups. Appendix A, Table A-2 (rows 2.2 and 2.3) shows the sum of weights before and after the nonresponse adjustment.

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

Variable	Levels
Sex of respondent	1. Male 2. Female
Presence of children and/or adolescents in the household at the screener level	1. Yes 2. No
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, 2001 California Health Interview Survey.

4.3 Adult Trimming Factors

Before going on to the last step of weighting, 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. If observations with large weights are found, the weights for these cases may be reduced in a process we call trimming. Trimming reduces the weight and the influence of the observation on the estimates and variances.

To identify influential weights that were candidates for trimming we examined the distribution of the weights by inspecting graphs of the weights. In addition, we computed two statistics, each a function of the weights, to help identify influential observations. The first statistic, RV_{D_k}, (computed for k = 0,1,2,3, and 4) measures each unit's effect on the variability of the weights (hence on the variance of the estimates). The D_k (i.e., D_0, D_1, …, D_4) defines the set of weights less the largest k weights. For example, D_1 is the set of all weights excluding the two largest weights. The statistic is

$$RV_{D_k} = \frac{USS_{D_k}}{(n-k)\,\bar{w}_{D_k}^2},$$

where $\bar{w}_{D_k} = \dfrac{n\bar{w} - \sum_{j \notin D_k} w_j}{n-k}$, $USS_{D_k} = USS - \sum_{j \in D_k} w_j^2$, w_j is the weight of the j^{th} observation, $\bar{w}$ is the mean of the weights, and USS is the uncorrected sum of squares of the weights.

A second statistic we considered looked at the spacings of the largest weights. The statistic is

$$d = \frac{z_n}{z_{n-1} + z_{n-2} + z_{n-3}}$$

where z_k is the space between the $y_{(k)}$ and $y_{(k-1)}$ order statistics defined as $z_k = y_{(k)} - y_{(k-1)}$. The values $y_{(n-4)}$, $y_{(n-3)}$, $y_{(n-2)}$, $y_{(n-1)}$, and $y_{(n)}$ represent the five largest order statistics for the adult weights. The two statistics, RV_{D_k} and d, and the distribution of the weights were examined separately by sampling strata. When either of the two statistics was large, the case was considered for trimming but not necessarily trimmed. The final decision on trimming also involved a manual inspection of the graphs.

The trimmed weight $TRMWT_i$ is

$$TRMW_i = TFACT_i \cdot ADA1W_i$$

where

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

where $0 < t_i < 1$.

Nine adult records were trimmed.[2] The trimming factor, t_i, ranged from 0.67 to 0.8. Table A-2 (row 3.1) shows the strata with trimmed weights, and rows 3.2 and 3.3 show the sum of weights before and after trimming.

4.4 Adult Raked Weight

The final step in the adult weighting was raking the trimmed adult weights to population control totals to produce estimates consistent with the Census 2000 results. The specific control totals and the approaches used to create them are described in Chapter 7. Raking is a commonly used estimation procedure in which estimates are controlled to marginal population totals. It can be thought of as a multidimensional poststratification procedure because the weights are poststratified to one set of control totals (a dimension), then these adjusted weights are poststratified to another dimension. The procedure continues until all dimensions are adjusted. The process is then iterated until the control totals for all the dimensions are simultaneously satisfied (at least within a specified tolerance). Raking is described more completely in Chapter 7.

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

$$RAKEDW_i = RAKEDF_k \cdot TRMW_i .$$

[2] 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 were examined again and new decisions were made about trimming. This might have changed the decision about which weights should be trimmed or the magnitude of the trimming factor. If this was done, 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 factor $RAKEDF_k$ is determined to satisfy the conditions

$$CNT_k = \sum_{i \in k} RAKEDF_k \cdot TRMW_i' ,$$

and CNT_k is the control total for raking dimension k. Table A-2 (rows 4.2 and 4.3) shows sum of weights before and after the raking adjustment.

4.5 Nontelephone Adjustments

Since CHIS is an RDD sample, households without telephones do not have a chance of being selected. An additional adjustment was implemented in CHIS 2001 to reduce biases from households without telephone numbers. The 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 complete interview did not posses the required data for the adjustment.

The decision to use a telephone survey introduces a coverage error due to the exclusion of persons in households without telephones. For estimates correlated with socioeconomic measures the associated bias may be important. A method for correcting this bias was suggested by Keeter (1995), who noted that the telephone status of a household changes over time and households with interruptions in telephone service are similar to households without telephones. Brick, Waksberg and Keeter (1996) took this idea and translated it into a weighting method to account for persons without telephones. The method works by adjusting the weights of sampled telephone households who have had telephone service interruptions.

This approach was considered for CHIS 2001, but it was not used due in part to the fact that very few records in CHIS 2001 data had interruptions in service (see Table 4-3). If we used the Keeter adjustment, the weights of the few records with interruptions would have been increased so much that they might have had undue influence on the final estimates.

Table 4-3. Number of adult completed screener interviews with interruptions in telephone service

Interruption of service	Count	Percent
No interruption	53,330	96.2
1 day	226	0.4
More than 1 day to 1 week	1,099	2.0
More than 1 week to 1 month	381	0.7
More than 1 month	355	0.6
Unknown	37	0.1
Total	55,428	100.0

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.

Instead of the Keeter method, the approach used in CHIS 2001 is a variation of a method proposed by Ferraro and Brick (2001). This method is an extension of the calibration method (Deville and Särndal, 1992) where the creation of calibration cells is done explicitly for the purpose of adjusting for households without a telephone. In Ferraro and Brick's method, logistic regression is used to compute the propensity to be a telephone household using an external file that includes households with and without telephones. Cells that are homogeneous with respect to the telephone status are formed by grouping cases with similar propensities. Using the same model, these cells are recreated in the sample file to be used as calibration cells. The external file can also be used to create the control totals since it includes both telephone and nontelephone households.

The Ferraro and Brick (2001) 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 specific procedures used for CHIS 2001 are now described. The March 2000 Current Population Survey (CPS) served as the external file for the nontelephone adjustment. The March 2000 CPS sample included telephone and nontelephone households and was large enough to produce reliable estimates for California. Table 4-4 shows the set of variables that are captured in both the CHIS and CPS. We tabulated various estimates to verify that both surveys produce consistent estimates. Only variables

that produced similar estimates[3] were considered as predictors. The estimates for the CHIS sample were computed using weights created for the July 2000 sample.[4]

Table 4-4. Common variables between CHIS 2001 and the March 2001 CPS

Variable for consideration	Variables used in cell creation	Variable description
Sex		Self-reported sex
Race/ethnicity	✓	Self-reported race and ethnicity
Education		Adult education level in household, based on high school education
Poverty		Poverty level, less than or greater than 100
Insurance		Insurance indicator, if anyone in household is insured
AFDC	✓	Household receiving aid from the AFDC program
Hhassist	✓	Household receiving public housing subsidies
Medicare		Household participates in MEDICAL program
Medicaid		Household participates in MEDICAID program
Typehh		Combination of number of adults, teens, and children in the household
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, 2001 California Health Interview Survey.

Some estimates of the CHIS 2001 variables did not mirror those of CPS as expected; thus they were not used. For example, we initially planned to use a poverty variable based on total household income to distinguish between households with and without telephones. However, the CHIS 2001 and March 2001 CPS questions used to create this poverty variable were asked in different ways, making the distributions of the variables dissimilar. In particular, CHIS 2001 asked for total household income in the past 12 months, while CPS asked for total household income the previous calendar year.

Some of the variables in the CHIS sample had missing values due to item nonresponse. These missing values were imputed using "hot-deck" imputation.[5] Hot-deck imputation is a technique where cases with missing values for specific variables are filled in with values from other cases. Potential donors (cases that may contribute a value) and recipients (cases with missing values) are classified into

[3] 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 in a direct way.

[4] 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 RDD sample and was used for preliminary analysis.

[5] "Hot-deck" imputation is discussed further in Chapter 8.

cells. The cells are constructed in such a way that characteristics are as homogeneous as possible for potential donors and recipients. Recipients are imputed values from donors within the same cell.

Once the set of predictors common to the CHIS and CPS files were determined, we created the calibration cells. The goal was to create cells where the households had a similar propensity of having a telephone. We used the categorical search algorithm CHAID (Kass, 1980) to divide the CPS data where the dependent variable was the telephone status (i.e., telephone household, nontelephone household). CHAID divides the data into groups so that the propensities between the cells are as different as possible. Given a set of categorical predictors, CHAID divides the data into groups in a stepwise fashion. Through a series of chi-square tests for equality of distributions, CHAID identifies the most important predictor 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 logistic regression as used in Ferraro and Brick (2001). First, the 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.

After the CHAID analysis, the same cells were created in the CHIS 2001 sample. Missing items for the variables used in the cell creation were imputed in the same way as described earlier. Since the CHIS 2001 sample was raked to control totals, the nontelephone adjustment was carried out as an additional dimension in the raking procedure. Table 4-5 shows the definition of the cells used for the nontelephone adjustment.

Table 4-5. Nontelephone adjustment cell definition for CHIS 2001

Person type	Cell	AFDC participant or receives public housing assistance	Number of children	Number of adults	Race/ethnicity
Adult	4	Yes	0 or 1		
Adult	5	Yes	2 or more		
Adult	10	No			Latino or Black non-Latino
Adult	11	No			Other non-Latino
Child	1	Yes	0 to 2		
Child	2	Yes	3 or more		
Child	6	No			Latino
Child	7	No			Non-Latino
Teen	3	Yes			
Teen	8	No		0 or 1	
Teen	9	No		2 or more	

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

5. CHILD WEIGHTING

A final child weight was created for all completed child extended interviews[1] in CHIS 2001. 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 responding adults because children are selected in the adult interview. A more 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 completed, the child and adolescent weights must be further adjusted to account for the nonresponse.

Children and adolescents were sampled in Section H of the adult extended interview. To account for adults that did not complete Section H of the adult interview (hence, no child or adolescent could be sampled), the household final weight $HHA5W_i$ was adjusted. We refer to this adjusted weight as the Section H adjusted household weight, $HHA6W_i$, and it is

$$HHA6W_i = HHA6F_c \cdot HHA5W_i ,$$

where

$$HHA6F_c = \begin{cases} \dfrac{\sum_{i \in SH\_C, SH\_NC} HHA5W_i}{\sum_{i \in SH\_C} HHA5W_i} & i \in SH\_C \\ 0 & i \in SH\_NC \end{cases} ,$$

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

and *SH_C* is the set of adults who completed Section H, *SH_NC* is the set of adults who did not complete Section H, and *c* denotes the Section H adjustment cell. Note that this adjustment can be considered as a household adjustment on top of the poststratified household weight.

The Section H nonresponse adjustment cells were created within sampling strata 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.

5.2 Initial Child Weighting

The initial child weight is the product of the adjusted household weight and the probability of sampling the child within the household. In CHIS 2001, 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 either the sampled adult or the sampled adult's spouse. If the sampled adult and his spouse did not have a child associated with them, then no child was sampled even if there were children present in the household. See Report 1: Sample Design for information on the within-household person selection process.

Since the child sampling depends on the associations of children and adults within the household, these associations were defined before sampling children. Children who were not associated with any adult in the household were randomly linked to an adult. Randomly linked children were only sampled through the sampled adult (not the spouse of the adult).

The overall probability of selecting a child within a household is thus the product of the probability of selecting the sampled adult and the probability of selection of the sampled child among the associated children. The initial child weight, $CHA0W_i$, is the product of the household weight and the inverse of the within household probability of selecting given by

$$CHA0W_i = \frac{1}{CHPROB_i} \cdot \frac{1}{ADCMP_i} \cdot HHA6W_i,$$

where $CHPROB_i$ is the probability of sampling child i, and $ADCMP_i$ is the probability of sampling the sample adult and his or her spouse (the adult component of the within household child weight). These two factors are described more completely below.

The probability of selection of the sampled child, $CHPROB_i$, is

$$CHPROB_i = \frac{1}{CHILDCNT_i},$$

where $CHILDCNT_i$ is the number of children associated with the sampled adult. The adult component of the child weight, $ADCMP_i$ is

$$ADCMP_i = \begin{cases} \dfrac{ADMOS_i + SP\_ADMOS_i}{\sum_i ADMOS_i} & \text{if the sampled adult is married and the sampled child is not randomly linked to the sampled adult} \\ \dfrac{ADMOS_i}{\sum_i ADMOS_i} & \text{otherwise} \end{cases}$$

where $ADMOS_i$ is the measure of size of adult *i* in the household (see Section 4.1), and SP_ADMOS_i is the measure of size of the spouse of the adult. The number of sampled children and sum of the initial weights are in Table A-3 (rows 1.1 and 1.2).

5.3 Other Child Weighting Adjustments

As mentioned before, the adjustments made to the child weights are the same as the adjustments done to the adult weights. These adjustments are for extended interview nonresponse, trimming influential weights, and raking to control totals. The raking adjustment included a dimension that accounted 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 the three variables: household mail status, sex, and child age group (0-3, 4-7, and 8-11 years old) within sampling strata. 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 mail status and sex within sampling strata at the state level. Any cells still containing fewer than 30 respondents were collapsed across mail status. We retained the cells defined by sex even though sex has only a small effect on child response rates because of the importance of sex and health characteristics. Table A-3 (rows 2.1 to 2.3) shows the number of sample records and sum of weights before and after the nonresponse adjustments.

The next step was to identify and trim 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 24 weights in the child component of CHIS 2001. The trimming factors range from 0.40 to 0.80. Table A-3 (rows 3.1 through 3.3) shows the distribution of trimmed weights by self-reported strata 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 Census 2000 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 A-3 (rows 4.1 through 4.3) shows the counts and sum of weights before and after the raking adjustments.

6. ADOLESCENT WEIGHTING

In CHIS 2001 adolescents were sampled and responded to the interview for themselves after a parent gave the interviewer permission to conduct the interview. Since adolescents are not a good source of information about their own insurance coverage, a separate interview was done with the parents of the adolescents on this topic. In this section we describe the creation of analytic weights for both the adolescent interview and the adolescent insurance interview.

6.1 Initial Adolescent and Adolescent Insurance Weights

Table 6-1 shows the cross-tabulation of the number of completed interviews for the adolescent and the adolescent insurance interviews. A much larger number of interviews with parents about insurance were completed than actual adolescent interviews. The reasons for this are discussed in Report 4: Response Rates.

Table 6-1. Number of completed adolescent and adolescent insurance interviews*

	Completed adolescent insurance		
Completed adolescent interview	Yes	No	Total
Yes	5,732	69	5,801
No	2,064		2,064
Total	7,796	69	7,865

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.

Because the response rates for the two types of interviews are so different, separate weights were created for analysis of the adolescent and adolescent insurance questionnaire data. The weight for analysis using data collected from the adolescent is referred to as "adolescent weight." The second weight is called the "adolescent insurance weight" and is appropriate for the analysis using data collected from the adolescent insurance interview only. For almost all analyses combining data from the two interviews, the adolescent weight is the most appropriate weight because it accounts for the higher nonresponse rate.

The procedures for creating the adolescent and adolescent insurance 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 both the adolescent and adolescent insurance 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, $TNA0W_i$, is

$$TNA0W_i = \frac{1}{TNPROB_i} \cdot \frac{1}{ADCMP_i} \cdot HHA6W_i$$

where $TNPROB_i$ is the probability of sampling adolescent *i*, and $ADCMP_i$ is the probability of sampling the sampled adult and his or her spouse (adult component of the within-household child weight). The probability of selection of the sampled adolescent, $TNPROB_i$, is

$$TNPROB_i = \frac{1}{TNCNT_i}$$

where $TNCNT_i$ is the number of adolescents associated with the sampled adult. The number of adolescents sampled is in Table A-4 (row 1.1). The sum of the initial adolescent and adolescent insurance weights are given in Table A-4 (row 1.2) and Table A-5 (row 1.2), respectively.

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. This step produced two adjusted weights, one for the adolescent interview and one for the adolescent insurance interview. Appendix A, Table A-5 shows the nonresponse adjusted adolescent insurance weight. Similarly, Appendix A, Table A-4 shows the nonresponse adjusted adolescent weight for the regular adolescent interview. For both weights, the adjustments were done in cells. To create the cells for each interview, we inspected response rates separately by mail status (mailable, nonmailable), sex (male and female), and age group (12-14 and 15-17 years old) at the state level. After reviewing these rates, we created cells using age groups and sex within sampling strata. Any cells containing fewer than 30 respondents were collapsed. Table A-4 (rows 2.1 to 2.3) shows the number of sample records and the sum of weights before and after the nonresponse adjustments for regular adolescent weights. Table A-5 (rows 2.1 to 2.3) shows the same factors for the adolescent insurance interview.

After the nonresponse adjustments, influential weights were identified and trimmed. Table 6-2 shows the number of records that were trimmed for both the adolescent and adolescent insurance weights. Table A-4 (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. Table A-5 (rows 3.1 to 3.3) gives the same data for the adolescent insurance weights.

Table 6-2. Number of trimmed weights and range of trimming factors for the adolescent and adolescent insurance weights*

Component	Number of weights trimmed	Range of trimming factors
Adolescent	20	0.33 to 0.80
Adolescent Insurance	17	0.33 to 0.80

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.

The last step of weighting was to rake the weights to the control totals. At this point there is a difference between the two sets of weights. We first describe the adolescent insurance raking and then the regular adolescent raking. The adolescent insurance weights were raked to population control totals to produce estimates consistent with the Census 2000 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.

The adolescent weights were also raked, but the control totals were derived from the raked adolescent insurance weights. This is known as sample-based raking, where a sample is used to benchmark a smaller sample. The procedure is called sample weighting adjustment by Brick and Kalton (1996) and Lundström and Särndal (1999) refer to this as Info-S calibration. We used sample-based raking to ensure that estimates from the adolescent and adolescent insurance interview are as consistent as possible. Table A-4 (rows 4.1 to 4.3) shows the number of sample records and sums of weights before and after the raking adjustment for regular adolescent weights, and Table A-5 (rows 4.1 to 4.3) do the same for adolescent insurance weights.

7. RAKING AND CONTROL TOTALS

This chapter describes the raking procedure and the development of control totals for the RDD and geographic samples. The first section gives a general overview of raking and why this procedure was used in CHIS 2001. The second section describes the 11 dimensions used in the CHIS 2001 raking for each type of weight. Ten of the dimensions use the demographic variables sex, age, geography, race, and ethnicity. The 11th dimension was created to reduce the bias associated with households without a telephone. The third section describes how the control totals for the dimensions were derived from the 2000 Census files.

7.1 Raking Procedure

Raking is an estimation procedure in which estimates are controlled to marginal population totals. In CHIS 2001, the adjustment to population control totals at the person level uses a raking procedure so that more auxiliary information could be included. For example, if poststratification were used only some age/race/sex categories could be used in the adjustments, while with raking more levels of these variables and important geographic level data such as county can also be included. As we mentioned earlier, 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). Below, we describe the procedure in more detail. Brackstone and Rao (1979) and Deville and Särndal (1992) also describe some aspects of raking.

Suppose we have two (we limit discussion to two for simplicity, but 11 are used in CHIS 2001) auxiliary variables with H and J classes, respectively. If we cross-classify the variables and the sample counts in some cells are small, then poststratification produces unstable estimates unless the cells in the cross-tabulation are collapsed. With 11 dimensions, the level of collapsing would have to be very extensive. An alternative estimation approach is to rake the weights to the marginal totals of the counts rather than the cell counts used in poststratification. The raking estimator is design-unbiased in large enough 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 weight can be written as $\tilde{w}_{cd,i} = w_{cd}\hat{a}_c\hat{b}_d$, where w_{cd} is the pre-raked weight of an observation in cell (*c*,*d*) of the cross-tabulation, $\hat{a}_c$ is the effect of the first variable, and $\hat{b}_d$ is the effect of the second variable. Note that in this formulation there is no interaction effect. In this sense, the weights are determined by the marginal distributions of the control variables. As a result, the sample sizes of the marginal distributions are the important determinants of the stability of the weighting procedure, not the cells formed by the crossing of the variables. This means that deficient cells (cells with small sample sizes) are defined by looking at the sample sizes of the margins. Furthermore, this permits the use of more variables or control totals with raking than is possible with poststratification.

7.2 Raking Dimensions

The 11 dimensions used in CHIS 2001 are shown in Table 7-1. The first 10 dimensions in Table 7-1 are created by combining demographic variables (age, sex, race, and ethnicity) and different geographic areas (city, county, group of counties, and state). The 11th dimension is created to adjust the weights for households without a telephone number. Section 4.5 has more details on the nontelephone adjustment and the variables used to create the levels for this dimension.

The weights for completed adult interviews were raked separately from the child and adolescent interview weights. The child weights were raked with the adolescent insurance weights rather than the adolescent interview weights because there were more completed adolescent insurance interviews (Table 6-1). The child and adolescent insurance weights could not be raked separately because the control totals of several raking dimensions included both children and adolescents. The adolescent weights were raked to the sum of the raked adolescent insurance weights in the sample-based raking procedure described in Section 6.2.

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

Dimension	Level	Description	Categories	
1	Stratum	Large age groups (3) × sex (2)	11	Under 12 years, males
			12	Under 12 years, females
			21	12 to 17 years, males
			22	12 to 17 years, females
			31	18 years or older, males
			32	18 years or older, females
2	Stratum	Small age groups (9)	1	Under 5 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	American Indian/Alaska Native indicator (2) × large age groups (2)	11	Yes, under 18 years old
			12	Yes, 18 years or older
			21	No, under 18 years old
			22	No, 18 years or older
4	State	Asian indicator (2) × large age groups (2)	11	Yes, under 18 years old
			12	Yes, 18 years or older
			21	No, under 18 years old
			22	No, 18 years or older
5	Collapsed stratum	Latino indicator (2) × large age groups (2)	11	Yes, under 18 years old
			12	Yes, 18 years or older
			21	No, under 18 years old
			22	No, 18 years or older
6	Collapsed stratum	African American indicator (2) × large age groups (2)	11	Yes, under 18 years old
			12	Yes, 18 years or older
			21	No, under 18 years old
			22	No, 18 years or older
7	Collapsed stratum	White indicator (2) × large age groups (2)	11	Yes, under 18 years old
			12	Yes, 18 years or older
			21	No, under 18 years old
			22	No, 18 years or older

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

Dimension	Level	Description	Categories	
8	State	Other indicator (2) × large age groups (2)	11	Yes, under 18 years old
			12	Yes, 18 years or older
			21	No, under 18 years old
			22	No, 18 years or older
9	State	Pacific Islander or Native Hawaiian indicator (2) × large age groups (2)	11	Yes, under 18 years old
			12	Yes, 18 years or older
			21	No, under 18 years old
			22	No, 18 years or older
10	State	Small age groups (13) × sex (2)	11	0 to 3 years, male
			12	0 to 3 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 25 years, male
			62	18 to 25 years, female
			71	26 to 30 years, male
			72	26 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
11	State	Nontelephone adjustment cells		See Table 4-5

Source: UCLA Center for Health Policy Research, 2001 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 required than is typical for poststratification. In dimensions 5, 6, and 7, the collapsed cells were created by combining counties into larger geographic areas. Dimensions 3, 4, 8, and 9 were defined at the state level because there were too few respondents in many of the cells at lower geographic levels. When collapsing, we ensured that there was at least one marginal that preserved the number of persons in the county/self-reported stratum. In this way, the raked weights added up to the control total for each stratum.

As noted in Section 6.2, to attain consistency between estimates from the adolescent interview file and the adolescent insurance interview file, the weights for the adolescent interview were sample-based raked to estimates from the adolescent insurance interview file. This required some additional definitions of the sample-based raking totals. For example, the total for number of adolescents is included in dimensions 1, 2, 10, and 11, but the remaining dimensions (dimensions 3 to 9) do not include separate totals for adolescents. As a result, the sample-based control totals for the number of adolescents for dimensions 3 to 9 (e.g., self-reported as Asian for dimension 4 or self-reported as Latino in dimension 9) are computed from the raked insurance weights. This total is an estimate subject to sampling variability. The adolescent weights should have the same variability as the adolescent insurance weights because the adolescent insurance weights were used as control totals during the raking. The variability was reflected by raking each replicate weight[1] to each replicate sum of weights of the insurance sample used to produce the control total. The only difference in the variability for these margins occurs when there were fewer than 50 respondents in the adolescent file and these required additional collapsing.

Table 7-2 shows the overall adjustment factors for the adult, child, and two adolescent weights. The overall adjustment factor was computed as the ratio of the control total to the sum of weights before raking. This factor is, 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 is not straightforward in CHIS 2001 because the weights are already adjusted at the household level for households with and without individuals under 18 years old. The adjustment factors confound several factors such as reporting error and residual nonresponse error, but still may be used as a rough indicator of within-household coverage error. A factor greater than unity suggests undercoverage, and a factor less than unity suggests overcoverage (these are all relative measures).

[1] See Section 10.3 for the creation of replicate weights.

Table 7-2. Overall adjustment raking factors for adult, child, adolescent and adolescent insurance interviews

Characteristic	Adult	Child	Adolescent	Adolescent insurance
Total	1.043	1.018	1.029	1.020
Sex				
Male	1.073	1.011	1.027	1.026
Female	1.016	1.025	1.031	1.014
Age group				
Under 5 years		1.022		
6 – 11 years		1.014		
12 – 17 years			1.029	1.020
18-24 years	1.094			
25-29 years	1.168			
30-39 years	1.095			
40-49 years	1.004			
50-64 years	0.954			
65 years and over	1.024			
Race				
White alone	1.036	0.983	0.977	0.980
African American alone	1.173	1.086	1.098	1.060
American Indian/Alaska Native alone	0.262	0.269	0.315	0.315
Asian alone	1.137	1.123	1.079	1.103
Native Hawaiian, Pacific Islander alone	0.500	0.477	0.622	0.498
Other race alone	1.121	1.131	1.361	1.410
Two or more races	0.879	0.934	0.922	0.769
Ethnicity				
Latino	1.069	0.946	1.079	1.116
Non-Latino	1.033	1.083	0.995	0.958
Urbanicity				
Urban	1.047	1.019	1.036	1.026
Rural	1.023	1.012	0.999	0.994
Number of children in the household				
None	1.026		0.999	0.986
One	1.065	1.024	1.043	1.030
Two	1.085	1.006	1.061	1.066
Three or more	1.099	1.028	1.127	1.151

Table 7-2. Overall adjustment raking factors for adult, child, adolescent and adolescent insurance interviews (continued)

Characteristic	Adult	Child	Adolescent	Adolescent insurance
Number of adolescents in the household				
None	1.044	1.018		
One	1.041	1.015	1.018	1.013
Two	1.035	1.015	1.036	1.019
Three or more	1.035	1.050	1.056	1.051
Presence of children or adolescents in the household				
Yes	1.064	1.018	1.029	1.020
No	1.027			

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.

Table 7-2 shows that the adults, children, and adolescents undercoverage for totals is minimal. For adults, the adjustment factor for males is larger than for females, which is common in household surveys. The factors also indicate a potential for undercoverage of adults 25 to 29 years old. In contrast, the factors for adults 50 to 64 years old are less than unity.

The most striking factors are for persons who self-reported as American Indian/Alaska Native alone, and Native Hawaiian or Pacific Islander alone. These factors are much less than unity, suggesting the CHIS 2001 estimates before raking across adults, child, and adolescents are much higher than the corresponding Census 2000 totals. We are unable to identify specific reasons why CHIS 2001 estimated so many more persons of these races than found in the Census 2000.

7.3 Producing the Control Totals

The control totals used in the raking are derived from the Summary File 1 (SF1) from the 2000 Census released by the U.S. Census Bureau. The SF1 contains information referred to as the 100 percent data, which is compiled from the questions asked of all people in every housing unit. Population items included sex, age, race, ethnicity (Latino), household relationships, and group quarters. The SF1 includes population and housing characteristics for the total population, population totals for an extensive list of race and Latino groups, and population and housing characteristics for a limited list of race and

Latino groups. The SF1 data are counts based on a new race classification used by the Census Bureau in which a person could report multiple races. The race classification includes six groups: white, black or African American, American Indian/Alaska Native, Asian, Native Hawaiian or Pacific Islander, and other race. Since a person could report multiple races, the SF1 provided counts for each of the 63 possible race combinations a person could report.

The SF1 is structured as individual files for each state. Only the California data are used for CHIS 2001. The file structure includes, but it is not limited to the geographic entities such as state, county, county subdivision, place (or place part), Census tract, block group, and block. For CHIS 2001, totals are computed at the county (or group of counties) level for most strata. The totals for the cities are based on the geographic level that the Census defines as places.

One of the limitations of using the SF1 for the control totals is the inability to produce counts that exclude the group quarter population for some dimensions used in CHIS 2001. The eligible population for CHIS 2001 includes only persons in residential households (not including those in group quarters—housing units with nine or more unrelated persons). Institutionalized persons in group quarters are also excluded. These persons include those living in prisons, jails, juvenile detention facilities, psychiatric hospitals and residential treatment programs, and nursing homes for the disabled and aged, or in military barracks.

The group quarters population should be excluded from the counts in the SF1 when deriving control totals for CHIS 2001. The group quarter population represented 2.4 percent of the total population in California (see Table 7-3); as a result, approximately 820,000 persons must be removed from the overall California population counts from the SF1.

Table 7-3. Population in California

Type	Population	Percent
In group quarters	819,754	2.4
Not in group quarters	33,051,894	97.6
Total	33,871,648	100.0

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

The difficulty arises because the group quarter counts that can be generated from the SF1 file are limited. Only counts by the characteristics listed in Table 7-4 can be produced. The table shows the group quarter counts from the SF1 are only available for three age groups (less than 18, 18 to 64, and 65

years old or older). The second dimension (DIM2) for CHIS 2001 requires separate counts for nine age groups, as shown in Table 7-1. The required control totals are computed by subtracting the group quarter population from the total population for the three age groups.[2] The new total is then allocated among the eight DIM1 age groups following the distribution of the age groups in the total population in the stratum. For age group 9 (65 year old or older), an allocation is not necessary because the subtraction produces this total directly.

Table 7-4. SF1 available group quarter counts

Characteristics	Available counts
Age groups	Less than 18 years old 8 to 64 years old 65 years old or older
Sex	Male Female
Race	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	Hispanic or Latino White alone not Hispanic or Latino

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

A similar procedure was used to compute control totals for the first dimension (DIM1, large age groups by sex). In this case, the new totals were computed by subtracting the group quarters counts separately for males and females in the two age groups (under 18 years old and 18 years old or older) by stratum. The new total was then allocated following the distribution of males and females in the two DIM1 age groups (under 12 and 12 to 17 years old) by stratum. There was no allocation for the age group that included 18 years old or older because the subtraction produced the control total directly. The dimension 10 (DIM10) control totals were computed using the same procedures used for DIM1 and DIM2. The only difference is that the controls were computed at the state level.

[2] These computations are done at the stratum level.

The control totals for dimension 5 (DIM5) were computed by subtraction of the Latino population in group quarters from the total Latino population by age group (under 18 and 18 years old or older).

The control totals for dimensions 3, 4, and 6 to 9 (DIM3, DIM4, and DIM6 to DIM9) are more involved. These dimensions are the total number of persons who reported a race, whether alone or in combination with other races. For example, for Asian Americans (DIM4), the control total includes persons who reported being Asian alone, Asian and African American, Asian and white, Asian and American Indian/Alaska Native, etc. The overall total is obtained by adding the totals for all 16 combinations of Asian reported with any of the other five races. Totals for this way of classifying persons excluding the group quarter population cannot be computed by subtraction and allocation because the appropriate group quarter counts are not available from the SF1. As a result, the population excluding those in group quarters for these dimensions had to be estimated based on the assumptions described below.

A simple way to explain the procedure used to estimate the control totals for these dimensions is with an example. Table 7-5 shows the counts needed to compute the population 18 years old or older not residing in group quarters who self-reported Asian alone or in combination with other races in a given stratum (DIM4).

Table 7-5. Computing the control totals for race

In group quarters			Not in group quarters			
Asian alone	Two or more races	One race, not Asian	Asian alone	Two or more races	One race, not Asian	Total
A_{1g}	R_{2^+g}	O_{1g}	$A_{1\bar{g}}$	$R_{2^+\bar{g}}$	$O_{1\bar{g}}$	T

The notation used is defined as follows:

T = the total number of persons in the stratum;

A_+ = the total number of persons who self-reported as Asian alone or in combination with one or more races;

$A_1 = A_{1g} + A_{1\bar{g}}$ = the total number of persons who self-reported as Asian alone;

A_{1g} = the total number of persons who self-reported as Asian alone living in group quarters;

$A_{1\bar{g}}$ = the total number of persons who self-reported as Asian alone not living in group quarters;

R_{2^+g} = the total number of persons who self-reported one race, not Asian living in group quarters;

$R_{2^+\bar{g}}$ = the total number of persons who self-reported one race, not Asian not living in group quarters;

O_{1g} = the total number of persons who self-report one race, not Asian living in group quarters;

$O_{1\bar{g}}$ = the total number of persons who self-report one race, not Asian not living in group quarters; and

$O_1 = O_{1g} + O_{1\bar{g}}$ = the total number of persons who self-reported other race alone other than Asian.

The control total needed for DIM4 is the number of persons not living in group quarters who self-reported Asian alone or Asian in combination with other races. Using our notation, this is given by

$$A_{+\bar{g}} = A_{1\bar{g}} + p_{A\bar{g}} * R_{2^+\bar{g}},$$

where

p_{Ag} = the proportion of persons that reported Asian and at least one other race living in group quarters; and

$p_{A\bar{g}}$ = the proportion of persons that reported Asian and at least one other race not living in group quarters.

The control totals cannot be computed because the proportions given above are not available from the SF1. To deal with this limitation, we solve the equation

$$A_{1g} + p_{Ag} * R_{2^+g} + A_{1\bar{g}} + p_{A\bar{g}} * R_{2^+\bar{g}} = A_+$$

subject to the following constraints:

$$\begin{aligned} p_{Ag} * R_{2^+g} &\le O_{1g} \\ p_{A\bar{g}} * R_{2^+\bar{g}} &\le O_{1\bar{g}} \\ 0 \le p_{Ag} &\le 1 \\ 0 \le p_{A\bar{g}} &\le 1 \end{aligned}$$

This process splits the number of persons who reported two or more races into two mutually exclusive groups. The first group includes all persons who reported Asian combined with other races while the second includes all the persons who reported any races other than Asian. The first group is added to the number of people who reported Asian abne to obtain the number of people who reported Asian alone or Asian combined with one or more races.

The equation has two unknowns and has an infinite number of solutions for p_{Ag} and $p_{A\bar{g}}$. To solve the equation we imposed conditions so that

$$p_{Ag} = p_{A\bar{g}} = p_A = \frac{A_+ - A_1}{A_+} = \frac{A_2}{A_+}$$

where A_2 is the number of people who reported Asian combined with one or more races. In other words, the proportion of people who reported Asian combined with one or more races living or not living in group quarters was the same proportion of people who reported Asian combined with one or more races among the people who reported two or more races in the stratum. For some race groups, the number of persons who reported two or more races living in group quarters was very small and in some cases it was zero. If it was zero then p_{Ag} was set to zero. In cases where $p_{Ag} * R_{2^+g} > O_{1g}$ then p_{Ag} was set to one. In either case, the equation was solved. Table 7-6 presents the computed counts by group quarters.

Table 7-6. Number of people in group quarters by race alone or in combination with one or more races

Description	Living in group quarters	Not living in group quarters	Total
Self-reported white	506,522	20,984,451	21,490,973
Self-reported African American	141,946	2,371,095	2,513,041
Self-reported American Indian/Alaska Native	15,476	612,086	627,562
Self-reported Asian	60,827	4,094,858	4,155,685
Self-reported Native Hawaiian, Pacific Islander	4,183	217,275	221,458
Self-reported other race	126,611	6,449,014	6,575,625

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

The last dimension (DIM11) was used to adjust for households without a telephone. The description and rationale for the dimension is given in Section 4.5. The control totals for the dimension are derived from the March 2000 CPS file. The percentage of the population for the cells was computed separately by person type (i.e. adult, child, and adolescent). The computed total number of persons not living in group quarters for each person type was then applied to these percentages to produce the control total of the cell. Table 7-7 shows the percentages and control totals for DIM11.

Table 7-7. Dimension 11 control totals

Person type	Cell	Control total	CPS percentage
Adult	4	591,077	2.5
Adult	5	360,950	1.5
Adult	10	7,689,182	32.2
Adult	11	15,207,044	63.8
Child	1	293,099	4.7
Child	2	363,350	5.8
Child	6	2,358,280	37.7
Child	7	3,237,268	51.8
Teen	3	326,449	11.1
Teen	8	279,095	9.5
Teen	9	2,346,100	79.5

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

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 2001. Report 4: Response Rates discusses unit nonresponse in more detail. This chapter focuses on item nonresponse and the imputation of missing response for certain variables. Only the imputation of variables used in weighting are covered here.[1] The imputed values were needed in the last stages of the weighting process and only interviews that were considered completed units were 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 county, sex, age, and race-ethnicity) and reviews the different types of imputation techniques used to fill in the missing data. The three techniques employed are deterministic (nonrandom), allocation, and hot-deck imputation. Section 8.2 discusses the imputation of self-reported county of residence and the assignment of self-reported stratum. Section 8.3 reviews the imputation process for self-reported sex and age. A similar imputation process was used for both variables. The final section of the chapter covers imputation of race and ethnic ity.

8.1 Imputed Variables and Methods

Table 8-1 lists the variables imputed in CHIS 2001. Most of the variables in the table are from items in the adult, child, and adolescent extended interviews. As noted above, the level of missing data is relatively small.[2] The specific percentages of missing data are given later in the chapter. When the amount of missing data is small and 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.

---

[1] Westat was responsible for imputing all the variables necessary for the weighting process. The CHIS project team at UCLA will evaluate the need for imputing all other items and implement procedures for imputing them as needed.

[2] The only variable with appreciable missing data is race/ethnicity for the adolescent insurance interview. The race and ethnicity of the adolescent are not obtained in the insurance interview so any adolescent who did not respond (but the parent did respond to the insurance items) had to be imputed.

Table 8-1. Description of imputed variables

Variable name	Description	Extended interview items	Variable type
SRAGE	Self-reported age	AAGE, CAGE, TEENAGE	Demographic
SRSEX	Self-reported sex	AA3, CA1, TA3	Demographic
AH42	County of residence	AH42	Geographic
SRSTRATA2	Self-reported strata	AH42, AM7, AM8, AM9	Geographic
SRH	Self-reported Latino	AA4, CH1, TI1	Ethnicity
SRW	Self-reported white	AA5_6, TI2_6, CH3_6	Race
SRAA	Self-reported African American	AA5_5, TI2_5, CH3_5	Race
SRAS	Self-reported Asian	AA5_4 TI2_4, CH3_4	Race
SRAI	Self-reported American Indian/ Alaska Native	AA5_3, TI2_3, CH3_3	Race
SRPI	Self-reported Pacific Islander	AA5_1, AA5_2, TI2_1, TI2_2, CH3_1, CH3_2	Race
SRO	Self-reported Other race	AA5_7 TI2_7, CH3_7	Race

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Three different imputation procedures were used in CHIS 2001 to fill in the missing responses. A flag indicating if the response is imputed accompanies every value. The first imputation technique is deterministic or non-stochastic in nature. Deterministic imputation was used to fill in the missing items for self-reported county of residence (item AH42). These imputations required no randomization because other geographic data are available that can be used to determine the respondent's county of residence with a relatively high level of precision (although not exactly).

The second imputation technique is a completely random selection from the observed distribution. This method is only used 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 8.3 of this chapter.

The third technique is hot-deck imputation, discussed in Section 4.5 of this report. Hot-deck imputation was used to impute race and ethnicity in CHIS 2001. 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. In order to carry out hot-deck imputation for CHIS 2001, 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 household structure. The recipient is then randomly imputed the same ethnicity and/or race (depending on the items that need to be imputed) from one of the donors in the pool. Once a donor is used, it is removed from the pool of donors. As with the other methods, imputation flags are used to track imputed values. Details and examples of hot-deck imputation are given in Section 8.4.

8.2 Self-Reported County and Self-Reported Stratum

County of residence is a household-level variable that is included only in the adult extended interview. The adult's county of residence is assigned to every sampled child and adolescent in the same household. Of the 57,848 completed adult interviews in CHIS 2001, 4.7 percent were missing county of residency. Across all types of interviews (adult, child, and adolescent extended interviews), there were 4,702 total records with missing county of residence.

Table 8-2 shows the number and percentage of imputed self-reported county of residence values for the adult extended interview. County of residence was imputed using one of several other sources of data. The sampled adult's self-reported ZIP Code (item AM7) was the primary source for filling in missing county responses. If a respondent reported a ZIP Code, we replaced the missing county by the county corresponding to the reported ZIP Code. The ZIP Code information obtained from the telephone exchange[3] was used to assign the self-reported county for the few cases where both self-reported county and ZIP Code were missing. Often, this amounted to assigning the county from which the respondent was sampled as the missing value.

[3] Genesys provides up to six ZIP Codes for a sampled exchange, including the most likely ZIP Code. These ZIP Codes were compared to each other during the imputation of county and ZIP Code.

Table 8-2. Number and percentage of completed adult interviews with missing county

County name	Number imputed	Percentage
Alameda	134	6.29
Alpine	0	0.00
Amador	5	3.31
Butte	22	2.59
Calaveras	7	3.47
Colusa	2	1.45
Contra Costa	54	4.32
Del Norte	3	2.16
El Dorado	26	3.21
Fresno	45	4.16
Glenn	8	3.36
Humboldt	12	1.66
Imperial	42	5.28
Inyo	5	4.67
Kern	47	4.26
Kings	36	4.30
Lake	9	2.47
Lassen	20	9.66
Los Angeles	702	5.50
Madera	32	3.88
Marin	15	1.96
Mariposa	0	0.00
Mendocino	6	1.32
Merced	33	3.87
Modoc	1	2.38
Mono	0	0.00
Monterey	119	14.84
Napa	21	2.51
Nevada	32	4.88
Orange	84	3.00
Placer	27	3.48
Plumas	4	2.94
Riverside	69	4.88
Sacramento	30	2.19
San Benito	5	62.50
San Bernardino	100	6.30
San Diego	87	3.11
San Francisco	61	3.03
San Joaquin	54	4.96
San Luis Obispo	13	1.60
San Mateo	52	5.11
Santa Barbara	23	2.26
Santa Clara	139	7.69
Santa Cruz	48	6.03
Shasta	55	4.83

Table 8-2. Number and percentage of completed adult interviews with missing county (continued)

County name	Number imputed	Percentage
Sierra	0	0.00
Siskiyou	13	2.71
Solano	94	5.99
Sonoma	20	2.54
Stanislaus	77	9.48
Sutter	16	3.86
Tehama	5	1.07
Trinity	2	1.65
Tulare	51	6.15
Tuolumne	7	2.81
Ventura	37	3.56
Yolo	23	2.71
Yuba	14	3.51

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

* Includes RDD sample and geographic and race-ethnic supplemental samples.

Another source of data for imputing self-reported county were two specific geographic items from the adult interview (items AM8 and AM9[4]). If the adult reported these items, the responses were geo-coded and were used to make the required county assignments. See Report 3: Data Processing Procedures for more details on the methods of geo-coding these responses.

A new variable, called the self-reported stratum, is derived from the fully imputed self-reported county of residence, where self-reported stratum corresponds to the 41 groupings of county used at the time of sampling. Self-reported stratum may differ from the sampling stratum because the sampling stratum may be incorrect or the respondent may incorrectly report county. Table 8-3 shows the distribution of self-reported stratum compared to the stratum at the time of sampling. This table shows that the sampling stratum and self-reported stratum are the same most of the time. The cities of Los Angeles county are the only exception and are discussed in more detail in Report 3: Data Processing Procedures.

The adult's self-reported stratum was derived from the self-reported (imputed when missing) county values for all strata except those strata corresponding to the cities in Los Angeles County (Pasadena and Long Beach) and Alameda County (Berkeley). For the city strata, the self-reported stratum

[4] These two items (AM8: What is the name of the street that you live on? and AM9: What is the name of the street down the corner from you that crosses your street?) were asked only in Los Angeles and San Diego.

Table 8-3. Distribution of self-reported strata versus sampling strata

Strata	Sampling	Self-reported	Net agreement ratio
Long Beach city	819	913	0.90
Pasadena city	814	671	1.21
Remainder of Los Angeles	10,582	10,612	1.00
San Diego	2,666	2,672	1.00
Orange	2,495	2,454	1.02
Santa Clara	1,514	1,508	1.00
San Bernardino	1,547	1,554	1.00
Riverside	1,386	1,391	1.00
Berkeley city	794	809	0.98
Remainder of Alameda	1,191	1,165	1.02
Sacramento	1,238	1,230	1.01
Contra Costa	1,199	1,214	0.99
Fresno	1,041	1,053	0.99
San Francisco	893	886	1.01
Ventura	971	1,015	0.96
San Mateo	925	945	0.98
Kern	1,096	1,093	1.00
San Joaquin	1,052	1,058	0.99
Sonoma	771	776	0.99
Stanislaus	819	794	1.03
Santa Barbara	798	795	1.00
Solano	1,587	1,553	1.02
Tulare	827	826	1.00
Santa Cruz	793	791	1.00
Marin	750	752	1.00
San Luis Obispo	799	807	0.99
Placer	784	764	1.03
Merced	832	849	0.98
Butte	825	835	0.99
Shasta	826	827	1.00
Yolo	834	844	0.99
El Dorado	780	807	0.97
Imperial	798	794	1.01
Napa	806	833	0.97
Kings	843	837	1.01
Madera	824	820	1.00
Monterey, San Benito	790	794	0.99
Del Norte, Humboldt	861	855	1.01
Lassen, Modoc, Siskiyou, Trinity	846	841	1.01
Lake, Mendocino	813	808	1.01
Colusa, Glen, Tehama	839	839	1.00
Sutter, Yuba	822	801	1.03
Plumas, Nevada, Sierra	814	824	0.99
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	818	813	1.01

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

* Includes RDD sample and geographic and race-ethnic supplemental samples.

assignment is based on self-reported ZIP Code. Since not all adults reported their ZIP Code, this item was imputed prior to making the assignment. Available geo-coded data and the telephone exchange data were used in this imputation process.

8.3 Self-Reported Sex and Age

The percentages of cases where either sex or age was missing in CHIS 2001 are miniscule across all types of extended interviews (adult, child, and adolescent). Table 8-4 shows the counts of missing data for these two variables and the percentage missing.

Table 8-4. Number and percent of completed interviews with missing self-reported sex and age

Person type	Number completed	Number missing sex	Percentage missing sex	Number missing age	Percentage missing age
Adult	57,848	0	0.00	11	0.02
Child	13,276	5	0.04	8	0.06
Adolescent	8,244	13	0.16	15	0.18
Total	79,368*	18	0.02	34	0.04

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

* Includes RDD sample and geographic and race-ethnic supplemental samples.

The missing data for self-reported sex was imputed randomly, corresponding to the distribution of sex in the population. In particular, a (uniformly distributed) random number was drawn for each missing value and the sex was assigned to be female if the random number was less than 0.50. Otherwise, the sex was imputed to be male.

A similar procedure was used to impute the missing self-reported age values. For each type of interview, random draws were made from the observed distribution of self-reported age and the missing value was replaced by the age corresponding to the randomly selected number. 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.

8.4 Self-Reported Race and Ethnicity

While the procedures used to impute the missing values of self-reported residential county, sex, and age were relatively straightforward, self-reported race and ethnicity presented a greater challenge. Different imputation processes were considered before settling on 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 was the realization that the method does not account for households in which persons have more than one race and ethnicity.

The hot-deck imputation procedure discussed below deals with the diversity of race and ethnicity within household 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 missing data we deal with are only the race items that classify a person as white, black, Asian, American Indian/Alaska Native, Pacific Islander, or other. Also, these items are treated as either all reported or all missing. In a very few cases there were missing values for one of the races but not others, but 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 8-5 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.

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 one 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

Table 8-5. Number and percent of completed interviews with missing self-reported race and/or ethnicity*

Type of interview	Imputed race**		Imputed ethnicity	
	Count	Percentage	Count	Percentage
Adult	1,754	3.03	216	0.37
Child	691	5.20	55	0.41
Adolescent	349	6.01	43	0.74
Adolescent Insurance	2,530	32.43	2,229	28.57
Total imputed***	4,980	6.28	2,500	3.15

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes RDD sample and geographic and race-ethnic supplemental samples.

**At least one value of race was imputed.

***The total imputed count does not equal the sum of the adult, child, adolescent and adolescent insurance counts due to overlap between the adolescent and adolescent insurance imputations.

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. When it was possible, the donors and the recipients had the same sampling strata. For cases where the pool formed in this way had too few donors, sampling stratum were combined based on geographic and urbanicity considerations. Once a donor was used, they were 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 one or more of the sampled persons might have a reported race and ethnicity in other households. The various combinations of reporting, such as a reported ethnicity but not race, were also encountered. Separate hot deck runs were made to accommodate all of these situations. Table 8-6 shows the counts and percentages of imputed self-reported race and ethnicity by type of extended interview (adult, child, adolescent, and adolescent insurance).

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 8-6. Counts and percentages of imputed self-reported race and ethnicity*

	Person type									
	Adult		Child		Adolescent		Adolescent insurance		Total**	
	Count	%	Count	%	Count	%	Count	%	Count	%
Self-reported race										
White alone	757	1.31	268	2.02	111	1.91	1,329	17.04	2,356	2.97
African American alone	31	0.05	8	0.06	7	0.12	139	1.78	178	0.22
Asian alone	41	0.07	7	0.05	3	0.05	208	2.67	256	0.32
American Indian/ Alaska Native alone	75	0.13	17	0.13	31	0.53	123	1.58	215	0.27
Pacific Islander alone	7	0.01	2	0.02	4	0.07	24	0.31	33	0.04
Other race alone	782	1.35	360	2.71	179	3.08	486	6.23	1,631	2.05
Two or more races	61	0.11	29	0.22	14	0.24	221	2.83	311	0.39
Self Reported Ethnicity										
Latino	58	0.10	21	0.16	18	0.31	701	8.99	780	0.98
Non-Latino	158	0.27	34	0.26	25	0.43	1,528	19.59	1,720	2.17
Completed interviews	57,848	100.00	13,276	100.00	5,803	100.00	7,801	100.00	79,368	100.00

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

* Includes RDD sample and geographic and race-ethnic supplemental samples.

**Counts in the Total column do not equal the sums of the rows due to overlap between the adolescent and adolescent insurance imputations.

9. WEIGHTING ADJUSTMENTS FOR THE RACE-ETHNIC SUPPLEMENTAL SAMPLES

This chapter describes the weighting adjustments and estimation issues for the race-ethnic supplemental samples, including Cambodian, Japanese, Korean, South Asian, Vietnamese, American Indian/Alaska Native, and Shasta Latino samples. Both geographic and race-ethnic supplemental samples were selected for CHIS 2001, but in this chapter we concentrate on the weighting process for the race-ethnic supplemental samples. The weighting for the geographic supplemental samples is included with the RDD sample weighting. Because of the special features of the supplemental samples, the data collected in these samples, estimation procedures and analytical weights for these groups are identified separately from the data and weights of the regular RDD and geographic supplemental samples.

To create the weights for the supplements for the race and ethnic groups the RDD and list samples were combined and weighted as a single sample. Although in previous chapters supplemental samples have been referred to as the records drawn from the special lists, in this chapter the term *supplemental sample* is used to describe the sample cases from both the RDD and list samples that meet the eligibility criteria for the specific race-ethnic group. The term *list sample* is used to describe sample cases drawn form the lists only, in contrast to sample cases drawn from the RDD sample. The weights for the supplemental samples are different from the RDD weights (described in previous chapters) because they are the result of a different weighting process discussed in the following sections.

This chapter is divided into six sections. The first section summarizes the alternatives considered before drawing the supplemental samples and indicates why the supplemental samples were selected using lists. The second section describes the two weighting approaches implemented. The first approach used a design-based weight that incorporates the probability of selecting the telephone number from the RDD and list frames. The second approach used a model-based weight that ignores the different probabilities of selecting the samples from the RDD and list frames. The third section discusses the creation of base weights. Weighting adjustment factors and raking for both approaches are described in the fourth and fifth sections. The last section evaluates the two weighting approaches and recommends which approach should be used to produce estimates for each supplemental sample.

9.1 Sampling Options

The race and ethnic subgroups are important for analytic reasons but constitute a small proportion of California's total population and are not highly concentrated in one or two areas within the state. As a result, the expected sample yield in a survey even as large as CHIS is too small to support making inferences for these subgroups at the desired level of precision. Because the members of these groups are a small percentage of the total population and are geographically dispersed, and no single list of all the members of the group is available, sampling methods for rare populations such as those described by Kalton and Anderson (1986) were considered for sampling these persons for CHIS.

Several sampling strategies were considered to increase the sample. These sampling strategies included household screening, the use of auxiliary information to classify and oversample telephone numbers, network sampling, and the use of special lists. Screening works by sampling a large number of telephone numbers, retaining a household if it contains a member of the rarest subgroup, and subsampling otherwise. A second strategy uses auxiliary information to stratify telephone exchanges by the proportion of members of the groups residing in these exchanges and then samples the strata at differential rates. A third sampling strategy is multiplicity or network sampling, where each household sampled that belongs to the targeted group is asked to identify other households linked to them (by linkage rules such as siblings). The linked households are then interviewed. A fourth scheme is a dual-frame design, in which the RDD sample is supplemented with a sample selected from a list of telephone numbers expected to include a high proportion of subgroup members.

The dual frame or supplemental list sampling method was chosen for CHIS 2001. The screening approach was too costly. The stratification method could not be implemented because the necessary auxiliary data were not available. The costs and yields for the multiplicity sampling approach could not be estimated in advance, and the measurement and nonresponse problems could not be tested in the time available before fielding the sample. Thus, the dual frame approach using lists was deemed to be the one most likely to succeed within the time and cost constraints of the survey.

In most dual frame surveys, the list does not contain all the members of the group, but the characteristics of the list, including its completeness, are very important. Of primary importance is that the list must contain the telephone number for members of the subgroup. A second important property is the completeness of the list to make the procedure more efficient. A third property is the need to cover a relatively broad spectrum of types of members of the groups so the efficiency for different types of

statistics for the subgroup can be improved. Finally, the list needs to be accurate in the sense of actually containing members of the group targeted; otherwise a large screening cost is incurred. *Report 1- CHIS 2001 Sample Design* contains additional details on the creation of the list frames and the selection of the supplemental samples from these lists.

9.2 Model-based and Design-based Approaches to Weighting

The estimation and statistical inference schemes used for the regular RDD and geographic supplemental samples are standard, design-based approaches appropriate for large probability samples. These data and accompanying weights described in previous chapters are appropriate for most CHIS analysis, including estimating characteristics of groups such as African Americans, Chinese, and Latinos that have relatively large sample sizes. Because of the special features of the race-ethnic supplemental samples, the data collected in the supplements and the estimation weights for the target groups were handled separately from the data and weights of the regular RDD sample. Special weights to produce estimates for members of these specific groups were created from these data.

Two approaches to estimating characteristics of the race-ethnic groups were considered and evaluated for each targeted group in CHIS 2001. One method is a design-based approach similar to that used for the regular RDD sample in that it assigns each household a base weight based on its probability of being included in the sample and makes standard nonresponse and population weighting adjustments to these base weights. These are *design-based* weights. The second estimation approach is referred to as a *model-based* approach, where each household was assigned a constant base weight and these base weights are then adjusted. Two sets of weights were created for the adult, child, and adolescent questionnaires.[1] These weights were created for the groups in Table 9-1. The creation of these weights is described in more detail below.

[1] The supplemental sample weights were not created for the adolescent insurance instrument.

Table 9-1. CHIS 2001 race-ethnic supplemental samples

Supplemental sample name	Description
AIAN sample	American Indian/Alaska Native sample from list from UCLA (primarily IHS)
South Asian*	South Asian sample from list of surnames from Genesys
Cambodian	Cambodian sample from list of surnames from Genesys
Japanese	Japanese sample from list of surnames from Genesys
Korean	Korean sample from list of surnames from Genesys
Vietnamese	Vietnamese sample from list of surnames from Genesys
Shasta County	Sample of Latino surnames in Shasta County from lists from Genesys

*South Asian includes Bangladeshi, Indian (India), Pakistani and Sri Lankan

Only persons who identified themselves solely or primarily as one of these race-ethnic groups[2] were eligible for interviewing in the supplemental samples. Because of this sampling restriction, the weighting procedures differed from those used in the RDD sample. When weighting the RDD sample (Chapters 2 to 7), the RDD weights were computed by multiplying the adjustment factors for the different stages of sampling and adjustment factors for nonresponse. In contrast, in the supplemental samples, the weights were computed by multiplying only the adjustment factors for the stages of sampling. Weights were adjusted for nonresponse during raking. In the last step, the weights were raked to control totals from the Census Bureau. Although the model-based weights did not reflect the probability of selection of the telephone number, they reflected most of the other weighting adjustments applied to the design-based weights, including raking. Some of these adjustment factors have been described in the previous chapters.

The weights for the supplemental sample can thus be written as a product of factors. The formula for each interview is presented and the factors are described in the subsequent sections. The adult final (raked) weight $ADOSWGT_i$ is

$$ADOSWGT_i = OSBSW_i * HHA1F_c * HHA4F_i * HHOSF_i * \left(\frac{1}{ADPROB_i} \right) * OSRKF_c$$

where

$OSBSW_i$ = the supplemental sample base weight;
$HHA1F_c$ = the subsampling adjustment factor
$HHA4F_i$ = the multiple telephone adjustment factor

[2] To be eligible for the AIAN oversample the persons must have responded that they were either AIAN-only, a member of a federally recognized tribe, OR were AIAN and some other race but most identified themselves as AIAN.

$HHOSF_i$ = the race-ethnic subsampling factor
$ADPROB_i$ = the probability of selection of the of the sampled adult
$OSRKF_c$ = the supplemental sample raking factor

The final supplemental sample child weight, $CHOSWGT_i$, is

$$CHOSWGT_i = OSBSW_i * HHA1F_c * HHA4F_i * HHOSF_i * \left(\frac{1}{ADCMP_i}\right) * \left(\frac{1}{CHPROB_i}\right) * OSRKF_c ,$$

and the adolescents, $TNOSWGT_i$, is

$$TNOSWGT_i = OSBSW_i * HHA1F_c * HHA4F_i * HHOSF_i * \left(\frac{1}{ADCMP_i}\right) * \left(\frac{1}{TNPROB_i}\right) * OSRKF_c ,$$

where

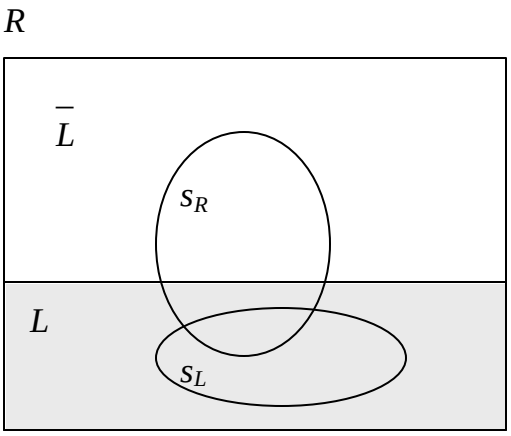
$ADCMP_i$ = the adult component
$CHPROB_i$ = the probability of selection of the sampled child
$TNPROB_i$ = the probability of selection of the sampled adolescent

Table C-1 in Appendix C describes the effect of the adjustments on the sums of weights and coefficients of variation for all the race-ethnic samples' model-based and design-based weights. The next two sections describe the base weights and the adjustment factors in the formula above for the model-based and design-based weights.

9.3 Creation of Base Weights

The design-based estimator is conceptualized by considering all telephone households in California as being in either on the supplemental list (*L*) or only as being eligible for sampling from the RDD sample (*R*) as shown in Figure 9-1. Figure 9-1 shows the relationship between the RDD and list frames and samples for a single list (i.e., Korean, Shasta, etc.) and a single stratum (i.e., county, city, or group of counties).

Figure 9-1. Dual frame



The notation in the figure are defined as:

R	the RDD frame containing the all telephone numbers.
N_R	the number of telephone numbers in the frame R.
s_R	the simple random sample drawn from the frame R.
n_R	the sample size (number of telephone numbers) in s_R drawn from the frame R.
L	the list frame (i.e. surnames, of telephone numbers in zip codes, etc).
N_L	the number of telephone numbers in the frame L.
s_L	the simple random sample drawn from the frame L.
n_R	the sample size (number of telephone numbers) in s_L drawn from the frame L.
$\bar{L}$	all the telephone numbers not found on the list.

We assume that all the numbers in the list are found in R, and $R = L \cup \bar{L}$. Note that the RDD sample s_R can be separated into two parts: s_{RL}, the portion of the RDD sample s_R that is found in the list L and $s_{R\bar{L}}$, the portion of the RDD sample 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, then sample $s_{R\bar{L}}$ is a simple random sample of size n_{RL} drawn from the $N_{\bar{L}}$ elements from stratum $\bar{L}$. Similarly, the sample s_{RL} is a simple random sample of size n_{RL} drawn from the N_L elements from stratum L. In strata L, there is a second sample s_L (the list sample). Since both samples s_L and s_{RL} are simple random samples, then we can treat them as a single sample of simple random sample of size $n_{RL} + n_L$ drawn from the N_L elements from stratum L.

The base weights are as follows:

For sampled records that could only be sampled from the RDD:

$$OSBSW_{\bar{L}} = \frac{N_{\bar{L}}}{n_{R\bar{L}}}$$

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

$$OSBSW_{L} = \frac{N_{L}}{n_{RL} + n_{L}}$$

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. The vendor of the lists (Genesys) provided these data and the complete AIAN list was available so the check could be made. 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, the weight produces an unbiased estimate in the traditional design-based framework.

Base weights were created for each RDD sampling stratum for all the race-ethnic groups except the Shasta Latino and the AIAN supplemental samples. The base weights for the Shasta Latino sample included only records from the Latino surname list and RDD records from the sampling stratum for Shasta County. For the AIAN supplemental sample the base weights were created for each sampling stratum separately by urban and rural areas.[3]

As mentioned before, in the model-based approach, the probability of selection of the sampled telephone numbers is ignored. In this case the base weight $OSBSW_i$ was defined as $OSBSW_i = 1$ for all the sampled telephone numbers independently of where the number was sampled.

Table G-1 (rows a through c) shows the number of sampled records, the sum of base weights, and coefficient of variation for the model and design approaches.

[3] Urban and rural areas defined by the U.S. Department of Indian Health Services (IHS)

9.4 Adjustment Factors

The base weights are then adjusted for other steps of sampling. As mentioned in Section 3.2, to reduce the number of calls to ineligible telephone numbers in household surveys (nonworking and business numbers) and improve the efficiency of the sample, telephone numbers in the RDD sample were subsampled at different rates based on their listed or mailable status (see Table 3-3). Listed and mailable telephone numbers drawn from the supplemental lists were not subsampled. The subsampling adjustment factor $HHA1F_c$ for the supplemental samples was computed as:

$$HHA1F_c = \begin{cases} 1 & \text{if } i \text{ was sampled from the list} \\ \dfrac{\sum_{i \in REVSTR=c} OSBW_i}{\sum_{i \in ISMP \text{ and } i \in REVSTR=c}} & \text{if } i \text{ was not sampled from the list and } i \in ISMP \text{ and } i \in REVSTR = c \\ 0 & \text{otherwise} \end{cases}$$

The sets *REVSTR* and *INSMP* are as defined in Section 3.2. The expression for the subsampling adjustment is the same for the model-based and design-based weights. Table C1 (Section 1.2) shows the number of sampled records, sum of weights, and coefficient of variation after this subsampling adjustment.

After adjusting the supplemental sample weights for subsampling, the weights were adjusted for the existence of multiple telephone numbers in the household (see Section 3.5). This adjustment was done only to the design-based weights. If the additional telephone number was used for residential purposes, then the household had a greater probability of selection since it could have been selected through the other number. The multiple telephone adjusted factor, $HHA4W_i$ was defined as

$$HHA4F_i = \begin{cases} 1 & \text{if the telephone number } i \text{ appears in the list frame or} \\ & \text{there is only one residential telephone number in the household} \\ 0.5 & \text{otherwise} \end{cases}$$

The underlying assumption is that there is a one-to-one correspondence between a telephone number and a household for all the numbers in the list. As a result, households represented by telephone numbers in the list do not have multiple chances of selection. The assumption is necessary because it is

not possible to determine the multiple chances of selection for numbers in the list. For telephone numbers not found in the list, this factor is similar to the multiple telephone adjustment factor for the RDD sample.

Under the model-based approach, all households have the same probability of selection. As a result, the multiple telephone adjustment factor $HHA4W_i$ was set to 1 for all the records in the combined RDD and list samples. Table C1 (rows 1.3 a. to c.) shows the number of sampled households where an adult extended interview was completed, sum of weights and coefficient of variation after the multiple telephone number adjustment for each race-ethnic supplemental sample.

After adjusting for multiple telephone numbers, the supplemental weights were adjusted for race-ethnic groups not interviewed in the supplemental samples. As mentioned before, the supplemental samples collected data only for persons who identified themselves primarily as members of these race-ethnic groups. In contrast, the RDD sample collected data for all persons independently of how they self-reported. To clarify this distinction, refer to Table 9-2. As an example, Table 9-2 shows the type of population interviewed in the RDD and list sample for Koreans.

Table 9-2. Race-ethnic group of sampled persons by type of interview

Sampled population		RDD sample	List sample
Respondent self-reported only one race or ethnic group (i.e., Korean)		A	*B*
Self-reports multiple races and/or ethnic groups (including Korean)	Self-reports as primarily Korean	C	*D*
	Self-reports as primarily not Korean	*E*	*F*

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

With households sampled from the list, only adults who identified themselves primarily as members of the race-ethnic group (i.e., Korean) were eligible and selected for the extended interview (cell *D* in Table 9-2). Households with adults that did not meet this criterion were coded as ineligible and the interview was terminated. Cell *F* in Table 9-2 represents the adults from these households. In contrast, if the household was selected from the RDD frame, one adult was always selected (cells *A* and *C*). Notice that adults who self-reported only one race ethnicity were always eligible for the extended interview independently of the frame from which they were sampled (cells *A* and *B*).

In the last step of the weighting adjustments, discussed in the next section, the weights were raked to the total number of adults in the race-ethnic groups. Since the population of interest for the supplemental samples included all persons who identified themselves primarily as members of the race-ethnic groups (cells *A*, *B*, *C*, and *D*), a suitable control total would include all persons who self-reported themselves in the same way. However, such totals are not available and could not be derived from the data collected in the Census 2000. The Census 2000 files include totals for the number of persons who self-reported only a given race-ethnic group or the number of persons who self-reported at least a given race-ethnic group and one or more others. If the total for the number of persons who self-reported only one group were used as a control for raking, the sample records for persons with more than one race-ethnic group could not be raked (cells *C*, *D*, and *E*). On the other hand, if the control total included all the persons who self-reported one or more race-ethnic group, then the records could not be raked because the list sample did not collect data for a part of the group (cell *F*).

To overcome this difficulty and properly rake the weights in the final step, the weights were adjusted to take into account the missing group not interviewed in the list sample. This is the race-ethnic adjustment, $HHOSF_i$, and it is

$$HHOSF_i = \begin{cases} 1 + \dfrac{\sum_{i \in B,D} HHA4W_i}{\sum_{i \in A,C} HHA4W_i} & \text{if the household } i \in E \\ 1 & \text{if the household } i \in A, B, C, \text{ and } D \end{cases}$$

where *A*, *B*, *C*, *D* and *E* are defined in Table 9-2. In this adjustment, the RDD records for households with adults who reported multiple races but who did not identify themselves primarily as the race-ethnic group of interest (cell *E*) are adjusted to represent themselves and similar households with adults not interviewed in the list sample (cell *F*). The underlying assumption is that the proportion of households with adults with multiple race-ethnic groups who do not identify themselves primarily as members of the group of interest is the same for the RDD and list samples. The adjustment for the AIAN sample was done using the same formula but separately for urban and rural areas. Furthermore, the expression of this adjustment is the same for the model-based and design-based weights. Table 9-3 shows the values of the race-ethnic adjustment factors for the supplemental samples for the design-based and model-based weights. Note that this factor is 1 for Shasta Latino because all adults in the list sample were interviewed even if the responding adult did not self-report himself primarily as Latino.

Table 9-3. Size of the race-ethnic adjustment factor

Supplemental sample	Race-ethnic factor	
	Design-based	Model-based
Cambodian	1.08	2.67
Japanese	1.34	1.67
Korean	1.41	1.65
South Asian	1.18	2.07
Vietnamese	2.05	2.67
AIAN (Urban)	1.97	1.69
AIAN (Rural)	1.26	1.37
Shasta Latino	1.00	1.00

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Table C-1 (rows 1.4 a to g) shows the number of sampled households where the adult extended interview was completed, sum of weights after the race-ethnic adjustment, the coefficient of variation, and the race ethnic factor for each supplemental sample.

The last adjustment factors prior to raking account for within-household sampling. For the creation of the adult weights, the adjusted household weights were adjusted to reflect the probability of selection of the sample adult. $ADPROB_i$ as described in Section 4.1. The difference for cases interviewed from the list sample is that only adults identified by the screener respondent as members of the target race-ethnic group were eligible for sampling. For the child weights, the household weight was adjusted to reflect the probability of selecting the sample adult and his or her spouse $ADCMP_i$ (i.e., the adult component of the within household child weight) and the probability of selection of the sampled child $CHPROB_i$. These two factors are described in Section 5.2. Similarly, the adolescent weights were created using the probability of selecting the sample adult and his or her spouse $ADCMP_i$ and the probability of selection of the sampled adolescent $TNPROB_i$ as described in Section 6.1. Table C-1 (rows in 2.1 for adults, 2.2 for children and 2.3 for adolescents) shows number of completed extended interviews, sum of weights, and coefficient of variation for the race-ethnic supplemental samples.

9.5 Raking Adjustment

The adult, child, and adolescent weights for the supplemental sample were raked to control totals in the last step of weighting. Raking is described in Section 7.1. Unlike the raking for the RDD

sample, the weights for completed adult, child, and adolescent interviews were raked together. The race-ethnic supplemental sample final (raked) weight $OSRKW_i$ is

$$OSRKW_i = OSRKF_k \cdot ADJW_i \, .$$

where $ADJW_i$ is the product of the base weight and all the previous factors and the factor $OSRKF_k$ is computed to satisfy the conditions

$$OSCNT_k = \sum_{i \in k} OSRKF_k \cdot ADJW_i \, ,$$

and $OSCNT_k$ is the control total for raking dimension k. Table 9-4 shows the control total dimensions used for the supplemental samples. Table C-1 (rows 3.1 for adults, 3.2 for children, and 3.3 for adolescents) shows the numbers of completed extended interviews, sum of weights after the raking adjustment and coefficient of variation for the race-ethnic supplemental samples.

The control totals used in the raking were derived from the Summary File 2 (SF2) from the 2000 Census released by the U.S. Census Bureau.[4] The SF2 contains information referred to as the 100 percent data compiled from questions asked of all people and about every housing unit. The SF2 includes population characteristics, such as sex by age, average household size, household type, relationship by household type, etc. The file provides information for 250 population groups: the population total, 132 race groups, 78 American Indian/Alaska Native tribes categories, and 39 Hispanic or Latino groups. The race groups in the SF2 data counts are based on a new race classification used by the Census Bureau in which a person could report multiple races (63 possible race combinations).

In addition to the 63 combinations of the main race groups, the SF2 file includes counts for specific Asian groups, including Cambodian, Vietnamese, Japanese, Sri Lankan, Pakistani, Bangladeshi, and Asian Indian (from India) reported alone or in combination with one or more races. These totals were used in raking for all the race-ethnic groups except for the supplemental sample for South Asians. The CHIS definition for South Asians comprises persons from India, Pakistan, Bangladesh, and Sri Lanka. The SF2 file does not provide a control total for the combination of these groups. The appropriate total would include all the persons who reported to be Indian, Pakistani, Bangladeshi, or Sri Lankan alone or in

[4] The weighting for the supplemental samples was delayed until the SF2 data for California were released. The SF1 data file was used for raking the RDD sample.

combination with one or more other races. To properly weight this group, an additional step was added in weighting, which is described in more detailed below.

As in the RDD weighting process, the group quarter population was excluded from the counts in the SF2 when deriving control totals (Section 7.3). The population in group quarters was subtracted from the total population. The new total was then allocated among the different levels of each dimension. After reallocating the new total, each dimension was examined to ensure that the totals were consistent across the levels of the dimensions. This process was done for all groups except the Shasta Latino and AIAN samples, for which the control totals were previously derived.

Each supplemental sample in CHIS 2001 had its own unique set of raking dimensions. These dimensions are shown in Table 9-4. Most of the supplemental samples had four dimensions; two dimensions created by combining age and sex, urban and rural areas, and number of persons who self-reported the race-ethnic group alone or combined with one or more other races. The Shasta Latino sample did not have an urban-rural dimension because that sample involved Shasta County only. The AIAN sample had sufficient records for a dimension that combined age groups and gender with the urban and rural areas. All respondents who self-reported the race-ethnic group either alone or combined were included in the raking process to produce a consistent estimate of the available control total as noted above.

As mentioned before, the SF2 files did not include an appropriate control total for the South Asian sample. Therefore, this supplemental sample required a slightly different approach. Before raking, the South Asian sample was used to create separate files for each ethnic groups of interest (i.e., Indian, Pakistani, Bangladeshi, and Sri Lankan). The SF2 file includes appropriate control total for each of these groups. These files were not mutually exclusive because a person who reported multiple ethnic groups could appear in different files (i.e., a person who self-reported as Indian and Pakistani was included in both the Indian and Pakistani files). Each of these files was then raked separately as for the other samples. After raking, the files were consolidated into a single file. For records with multiple weights, the final weight was computed as the average of the raked weights. In the example, if the record was in both the Indian and Pakistani files, the weight after raking each of these files was averaged and only one record was retained.

Table C-1 (row 3.4) shows the overall adjustment factors for the adult, child, and adolescent design-based weights. The overall adjustment factor is the ratio of the control total to the sum of weights

before raking. Unlike the RDD raking factor, the supplemental sample raking factor confounds a nonresponse adjustment factor and a measure of the magnitude of the bias correction for estimates of totals. Because the weights for the supplemental samples were not adjusted for nonresponse, the raking factor cannot be used even as an indirect measure of undercoverage.

Table 9-4. Dimensions used in raking

Oversample	Dimension	Description	Categories
Japanese	1	Age groups × Sex	1 Under 8 years, Males and Females 2 8 to 11 years, Males and Females 3 12 to 17 years, Males and Females 4 18 to 30, Males 5 18 to 30, Females 6 31 to 37, Males 7 31 to 37, Females 8 38 to 45, Males 9 38 to 45, Females 10 46 to 53, Males 11 46 to 53, Females 12 54 to 64, Males 13 54 to 64, Females 14 65 +, Males 15 65 +, Females
	2	Age groups x Sex	1 Under 18 years, Males 2 Under 18 years, Females 3 18 + years, Males and Females
	3	Race indicator	1 Japanese only 2 Japanese in combination with other races
	4	Urbanicity indicator	1 Urban 2 Rural

Table 9-4. Dimensions used in raking (continued)

Oversample	Dimension	Description	Categories
Korean	1	Age groups x Sex	1 Under 12 years, Males and Females 2 12 to 17 years, Males and Females 3 18 to 25 years, Males 4 18 to 25 years, Females 5 26 to 30 years, Males 6 26 to 30 years, Females 7 31 to 37 years, Males 8 31 to 37 years, Females 9 38 to 45 years, Males 10 38 to 45 years, Females 11 46 to 64 years, Males 12 46 to 64 years, Females 13 65 plus years, Males 14 65 plus years, Females
	2	Age groups x Sex	1 Under 7 years, Males 2 Under 7 years, Females 3 8 to 17 years, Males 4 8 to 17 years, Females 5 18 + years, Males and Females
	3	Race indicator	1 Korean only 2 Korean in combination with other races
	4	Urbanicity indicator	1 Urban 2 Rural

Table 9-4. Dimensions used in raking (continued)

Oversample	Dimension	Description	Categories
Vietnamese	1	Age groups x Sex	1 Under 12 years, Males 2 Under 12 years, Females 3 12 to 17 years, Males and Females 4 18 to 30 years, Males 5 18 to 30 years, Females 6 31 to 37 years, Males 7 31 to 37 years, Females 8 38 to 45 years, Males 9 38 to 45 years, Females 10 46 to 53 years, Males 11 46 to 53 years, Females 12 54 to 64 years, Males 13 54 to 64 years, Females 14 65 plus years, Males 15 65 plus years, Females
	2	Age groups x Sex	1 Under 8 years, Males 2 Under 8 years, Females 3 8 to 17 years, Males 4 8 to 17 years, Females 5 18 + years, Males and Females
	3	Race indicator	1 Vietnamese only 2 Vietnamese in combination with other races
	4	Urbanicity indicator	1 Urban 2 Rural

Table 9-4. Dimensions used in raking (continued)

Oversample	Dimension	Description	Categories
Cambodian	1	Age groups x Sex	1 Under 12 years, Males and Females 2 12 to 17 years, Males and Females 3 18 to 29 years, Males and Females 4 30 to 39 years, Males and Females 5 40 to 49 years, Males and Females 6 50 plus years, Males and Females
	2	Age groups x Sex	1 Under 18 years, Males 2 Under 18 years, Females 3 18 to 39 years, Males 4 18 to 39 years, Females 5 40 plus years, Males 6 40 plus years, Females
	3	Race indicator	1 Cambodian only 2 Cambodian in combination with other races
	4	Urbanicity indicator	1 Urban 2 Rural

Table 9-4. Dimensions used in raking (continued)

Oversample	Dimension	Description	Categories
AIAN	1	Age x Sex x Urbanicity	1 Under 12 years, Males, Urban 2 Under 12 years, Females, Urban 3 12 to 17 years, Males, Urban 4 12 to 17 years, Females, Urban 5 18 to 30 years, Males, Urban 6 18 to 30 years, FemMales, Urban 7 31 to 37 years, Males, Urban 8 31 to 37 years, FemMales, Urban 9 38 to 45 years, Males, Urban 10 38 to 45 years, Females, Urban 11 46 to 53 years, Males, Urban 12 46 to 53 years, Females, Urban 13 54 to 64 years, Males, Urban 14 54 to 64 years, Females, Urban 15 65 plus years, Males, Urban 16 65 plus years, Females, Urban 17 Under 12 years, Males, Rural 18 Under 12 years, Females, Rural 19 12 to 17 years, Males, Rural 20 12 to 17 years, Females, Rural 21 18 to 30 years, Males, Rural 22 18 to 30 years, Females, Rural 23 31 to 37 years, Males, Rural 24 31 to 37 years, Females, Rural 25 38 to 45 years, Males, Rural 26 38 to 45 years, Females, Rural 27 46 to 53 years, Males, Rural 28 46 to 53 years, Females, Rural 29 54 to 64 years, Males, Rural 30 54 to 64 years, Females, Rural 31 65 plus years, Males, Rural 32 65 plus years, Females, Rural
	2	Race x Urbanicity	1 AIAN Only, Urban 2 AIAN in combination with other races, Urban 3 AIAN Only, Rural 4 AIAN in combination with other races, Rural

Table 9-4. Dimensions used in raking (continued)

Oversample	Dimension	Description	Categories
Shasta County Latino	1	Age groups x Sex	1 Under 12 years, Males 2 Under 12 years, Females 3 12 to 17 years, Males and Females 4 18 to 37 years, Males 5 18 to 37 years, Females 6 38 plus years, Males 7 38 plus years, Females
	2	Age groups x Sex	1 Under 8 years, Males and Females 2 8 to 17 years, Males 3 8 to 17 years, Females 4 18 + years, Males and Females
	3	Race indicator	1 White only 2 Other
South Asian	1	Age groups x Sex	1 Under 4 years, Males 2 Under 4 years, Females 3 4 to 11 years, Males 4 4 to 11 years, Females 5 12 to 17 years, Males and Females 6 18 to 25 years, Males 7 18 to 25 years, Females 8 26 to 30 years, Males 9 26 to 30 years, Females 10 31 to 37 years, Males 11 31 to 37 years, Females 12 38 to 45 years, Males 13 38 to 45 years, Females 14 46 plus years, Males 15 46 plus years, Females
	2	Age groups x Sex	1 Under 8 years, Males and Females 2 8 to 17 years, Males 3 8 to 17 years, Females 4 18 + years, Males and Females
	3	Race indicator	1 South Asian only 2 South Asian in combination with other races
	4	Urbanicity indicator	1 Urban 2 Rural

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

9.6 Evaluation of the Weights

The reason both the design-based and model-based weights were computed is that neither is clearly the most appropriate for all the supplemental samples. The main problem with the design-based estimator is that the RDD and list-based weights may be very different, resulting in estimates with very low precision. Combining the RDD and supplemental samples increases the number of sampled cases; however, it does not necessarily increase the precision of the estimates. The model-based approach was developed to avoid the problem of having very different base weights by starting with the household base weights being equal for the RDD and list samples. The model-based estimates are not unbiased in the traditional sense, but they should have considerably lower variances than the ones computed using design-based weights.

One method used to quantify the variance of the estimates under the two approaches was to compute an approximate design effect (loss in efficiency due to differential weights) of a sample (Kish, 1992). The design effect for domain *d* (under conditions described by Kish) is approximately

$$DEFF_d = 1 + cv_d^2,$$

where *cv* is the coefficient of variation of the weights. Table C-2 in Appendix C shows the sample sizes, sums of weights, and coefficients of variation for the RDD, list, and combined RDD and list samples for the race-ethnic groups for the both model-based and design-based weights.

One reason the design effect is an important measure is that it relates the effective sample size, defined as

$$n_{eff} = \frac{n}{DEFF},$$

and the observed (or nominal) sample size, n. The effective sample size is an easily understood measure of precision of the estimates. Another useful statistic is the *DEFT,* which is the square root of the *DEFF*. The *DEFT* is useful because it is measured in the same units as the standard errors of the estimates.

Since the *DEFT* computed using the formula above is an approximation and some of the conditions required for it to be valid may be questioned, sampling errors for a number of statistics from the adult file were also computed using both the model-based and design-based weights. This

computation is discussed below. Table 9-5 summarizes all of these data across the supplemental samples for the adult file. The first column is the ratio of the *DEFT* of the design-based weight to model-based weight using the Kish approximation given above. The second column is a similar ratio of the design-based to the model-based estimates, but in this case it is of the average of the standard errors actually computed from the adult data file. The remaining columns give the total responding sample sizes and effective sample sizes using the design-based and model-based weights. These estimates are discussed later.

Table 9-5. Effective sample sizes for CHIS 2001 supplemental samples- adult interview

Oversample group	Ratio of design- to model-based DEFTs	Ratio of design- to model-based standard errors	Total sample*		
			Total number of respondents	Effective sample size (design-based)	Effective sample size (model-based)
Japanese	1.19	1.15	798	420	596
Korean	1.10	1.13	791	521	625
AIAN (Urban)	1.06	1.08	799	392	444
AIAN (Rural)	1.27	1.19	773	286	460
AIAN (All)	1.14	1.10	1,572	678	882
Vietnamese	1.10	1.08	843	471	567
Cambodian	1.70	1.87	197	50	144
Shasta	2.24	2.01	339	44	223
South Asian	1.38	1.27	822	331	626

* Race of the sample is defined as single race or most identified race if more than one race was reported.

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

As noted above, the main disadvantage with the model-based weights is the potential bias of the estimates. The bias of the model-based weights will be large if the persons on the list have characteristics that are different from the persons in the group that are not on the list. As part of the evaluation of the weights, the potential bias is examined by comparing characteristics of the sampled persons using the design-based and model-based weights. These differences are not reliable estimates of bias, but they do reveal whether the biases in the model-based estimates are likely to be very large. If the estimates with the two weights are very different, then the model-based approach may have too large biases to be useful.[5]

[5] It is also possible that a large ratio could arise because the design-based estimate is subject to high sampling variability.

To evaluate the potential bias from the model-based estimates, estimates of totals and percentages for 38 variables were computed from the adult extended questionnaire. The estimates are for adult characteristics only and are listed in Table C-3 in Appendix C. The selected variables include heath-related behaviors and socio-demographic characteristics, including at least one question from each extended interview section. A total of 118 different estimates and standard errors were produced for each weight (these are the standard errors shown in Table 9-5). In addition to these point estimates, the following statistics were computed:

- The difference between estimates produced using the design-based and model-based weights;
- The relative difference defined as the difference of the estimates with respect to the standard error of the estimate computed using the model-based weights; and
- The ratio of the standard error of the design-based estimate to the standard error of the model-based estimate (this is the number in column 2 of Table 9-5).

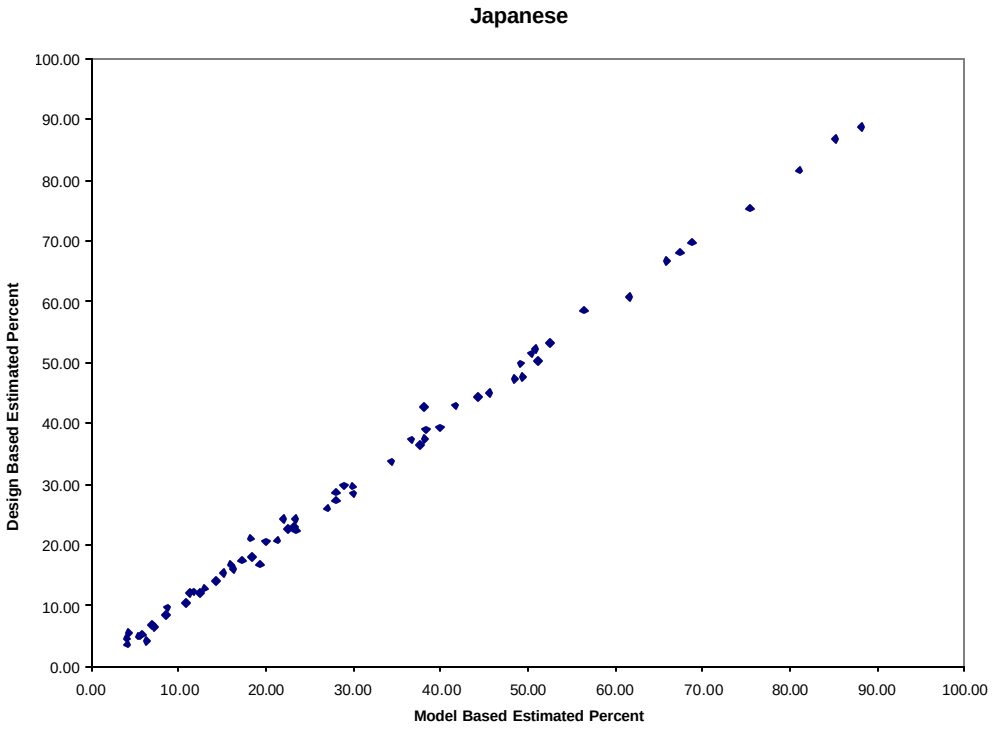
These statistics ratios can be used to approximate the overall distribution of the estimates, including the mean, median, minimum, maximum, 25th percentile, and 75th percentile. Summary statistics were computed for each supplemental sample of these distributions. However, rather than examine the tabular display, it was found that the implications were clearer when they were plotted. Figure 9-2 presents seven graphs of these statistics. For each supplemental sample, the upper graph is a scatter diagram of the estimates for a group computed using the model-based weights (on the horizontal scale) and the design-based weights (on the vertical scale). If the estimates fall in close proximity to the main diagonal, then the bias of the model-based estimates is not large. The lower graph shows a scatter diagram of the corresponding standard errors. In this case, points near the main diagonal indicate that the design-based standard errors are not that much larger than the model-based ones and the two weighting approaches produce estimates with about the same effective sample size.

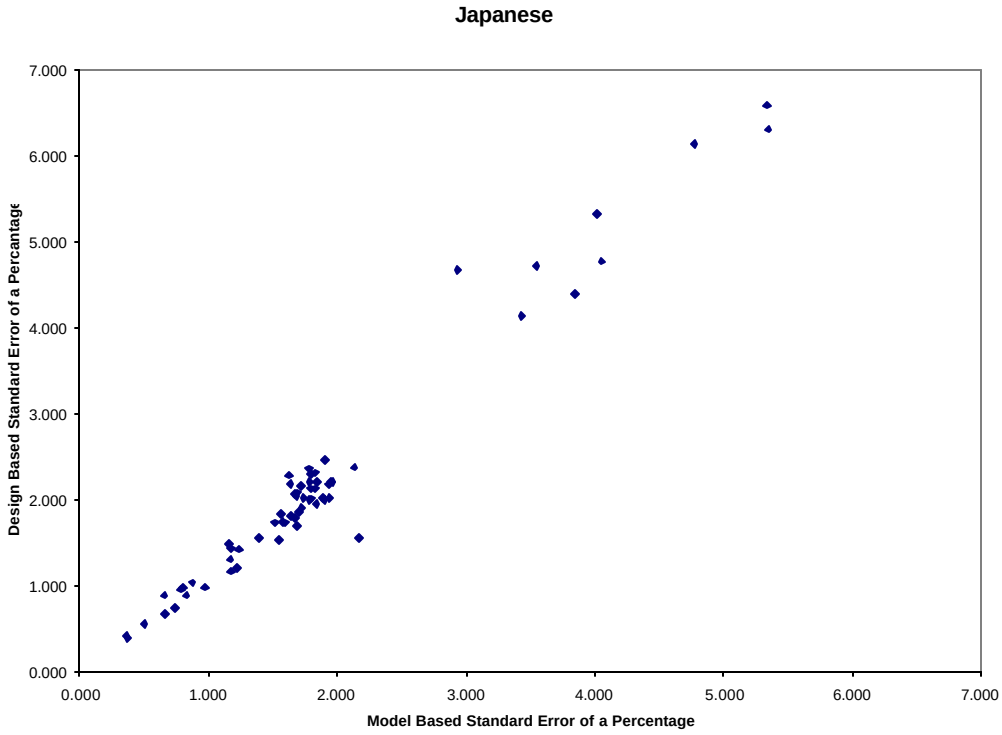
The first four graphs are for the Japanese, Korean, Vietnamese, and AIAN samples. The graphs are about the same for each of these samples. While the estimates for the model-based and design-based procedures give estimates that are similar, the standard errors are also not very different for the two methods. For example, for the Japanese sample the mean ratio of standard errors is 1.17. Whenever the standard errors for the design-based estimates are relatively close to those for the model-based estimates, the design-based procedure was chosen to reduce the potential for bias in the estimates. For example, the sizeable differences in educational attainment and marital status for the Japanese supplemental sample

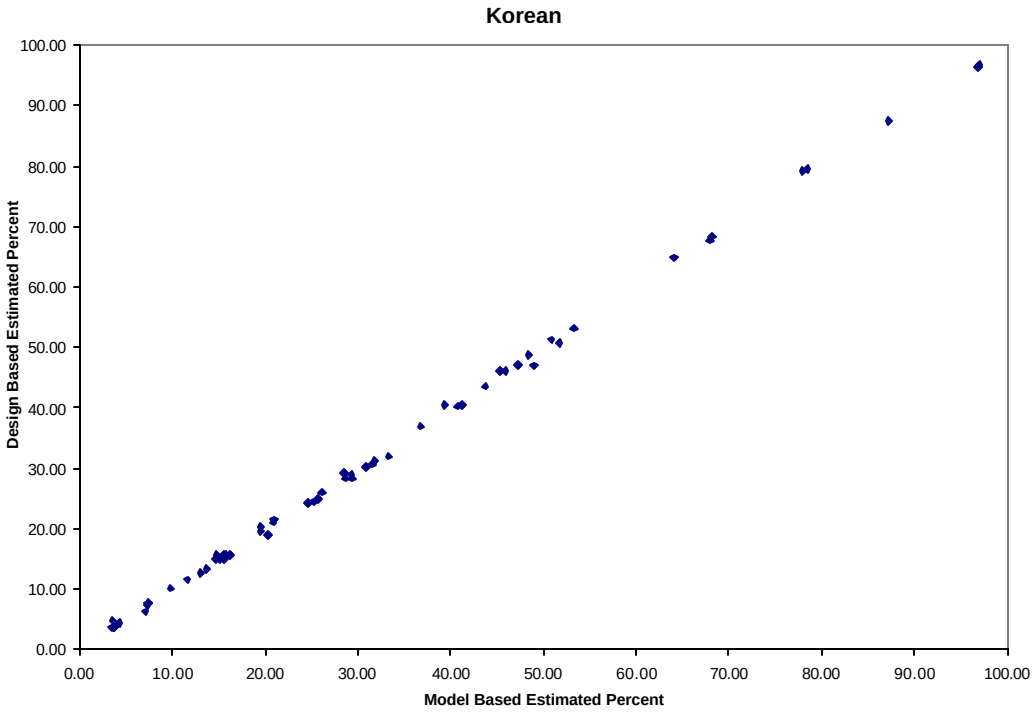
when the two approaches are used could be due to biases from the model-based weights. The more standard and defensible design-based weights were chosen in these cases.

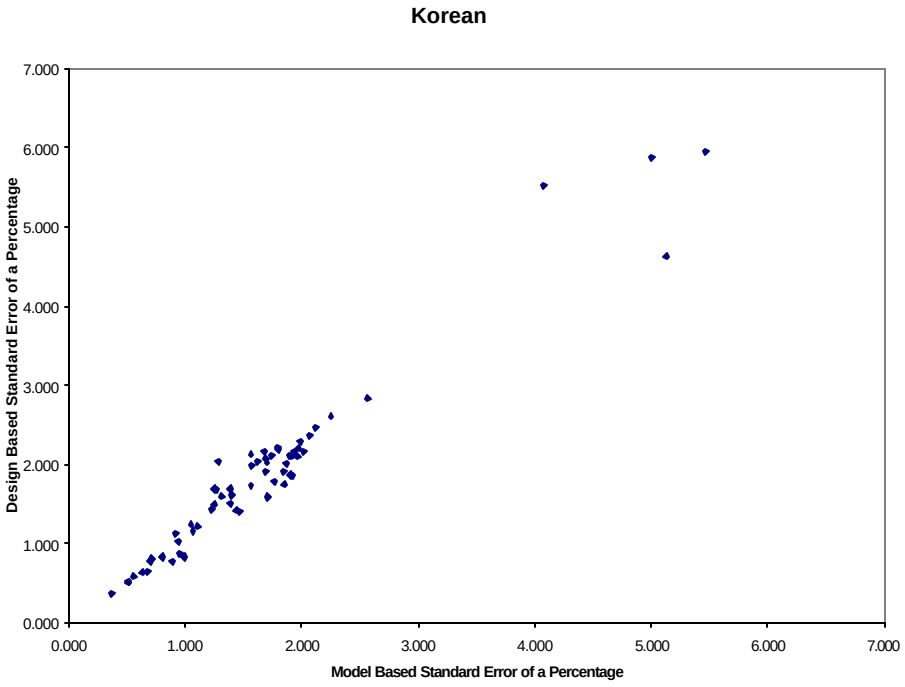
The design-based procedure was also selected for the South Asian sample, even though the model-based procedure gives smaller standard errors as shown in the plot. The effective sample size for the model-based estimates is approximately 626 while for the design-based effective sample size it is 331. The choice of the design-based weights for the South Asian sample is not as unambiguous because there is a substantial variance penalty with this choice. However, the design-based procedure is still preferred because of the potential for substantial biases in some model-based estimates.

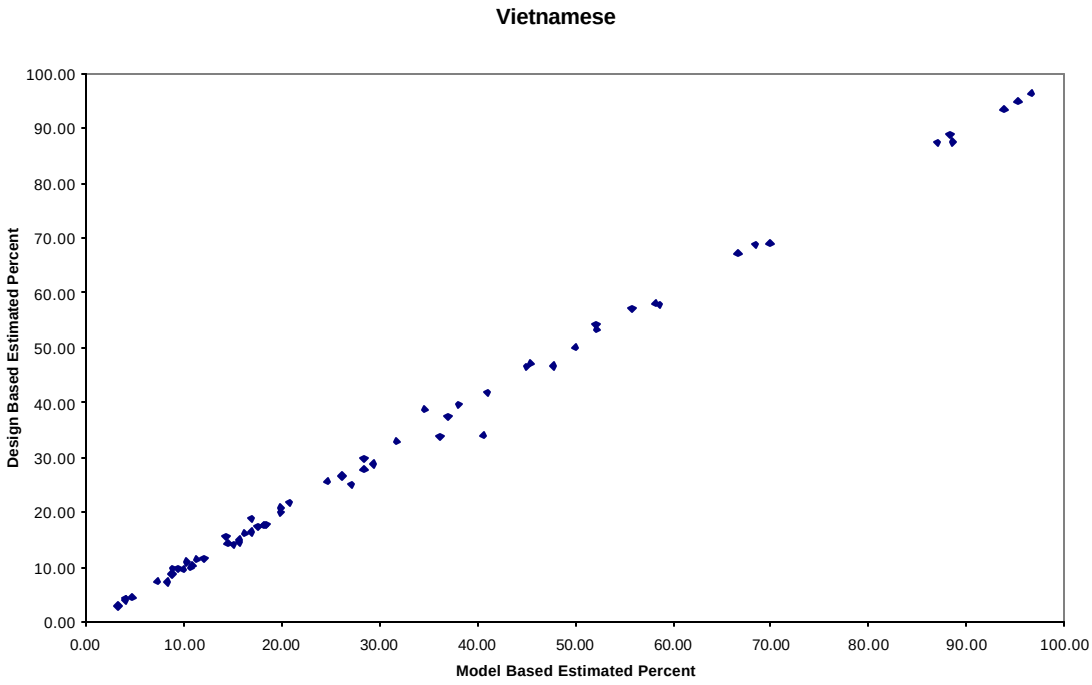
The last two groups, the Cambodian sample and the sample of Latinos in Shasta county, are very different from the previous samples. The effective sample size for the Cambodian model-based weights is 144 while for the design-based procedure the effective sample size is only 50. This difference is even more pronounced for the sample of Latinos in Shasta County, with a model-based effective sample size of 223 and a design-based effective sample size of 44. Even though the plots show some large differences between the model-based and design-based estimates, using the model-based weights may be beneficial because the sample size is very small with design-based methods. The model-based estimates appear to be the best choice in these two samples, but any analysis of the groups with the model-based weights should be cognizant of the possibility of large biases due to the composition of the list sample.

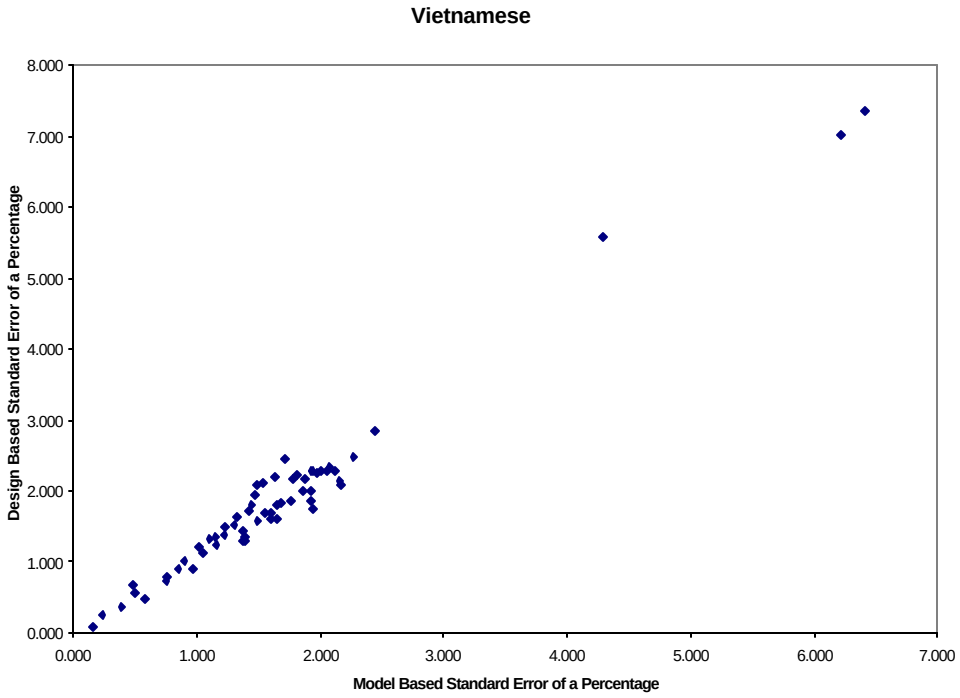

Japanese
Design Based Estimated Percent
Model Based Estimated Percent
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00

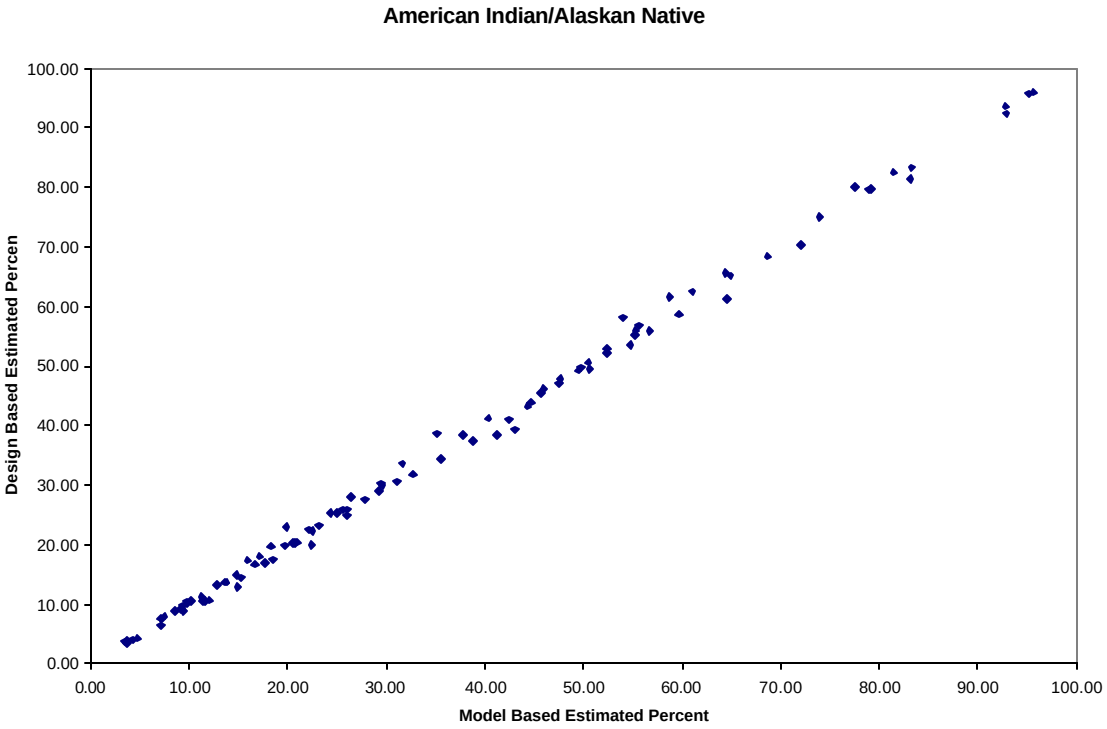

Japanese
Design Based Standard Error of a Percentage
Model Based Standard Error of a Percentage
0.000
1.000
2.000
3.000
4.000
5.000
6.000
7.000

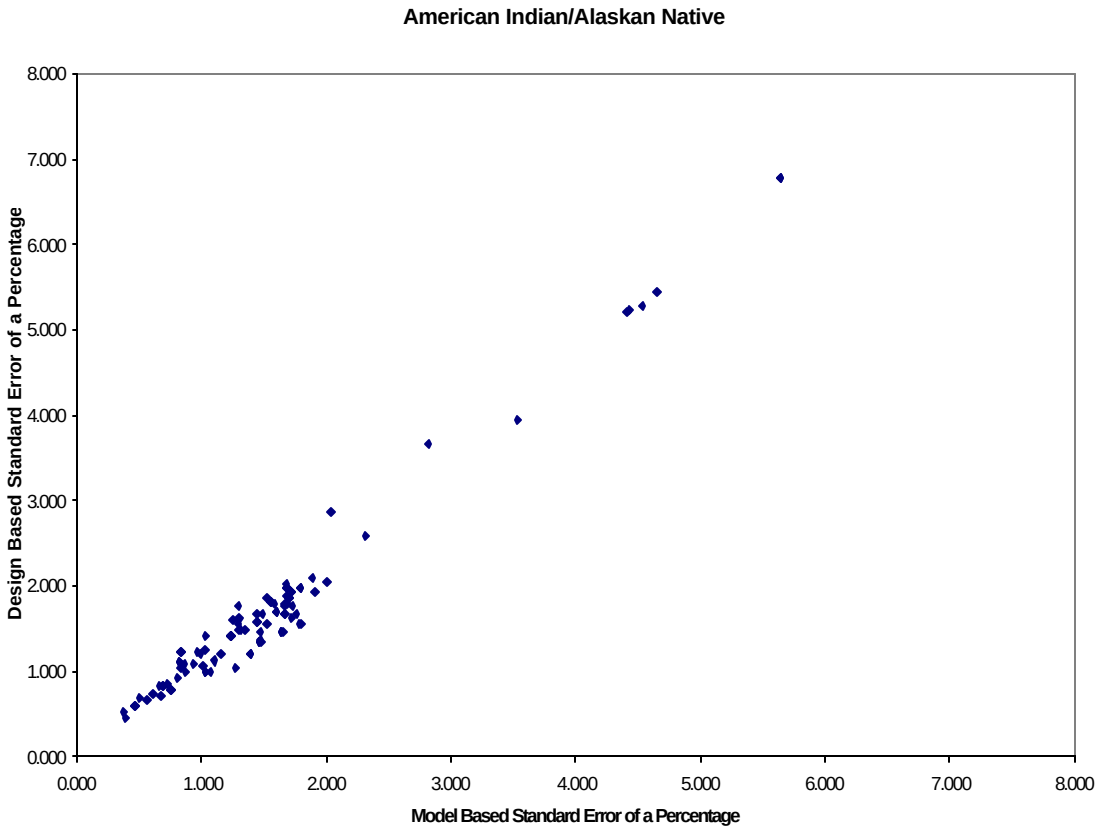

Korean
Design Based Estimated Percent
Model Based Estimated Percent
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00

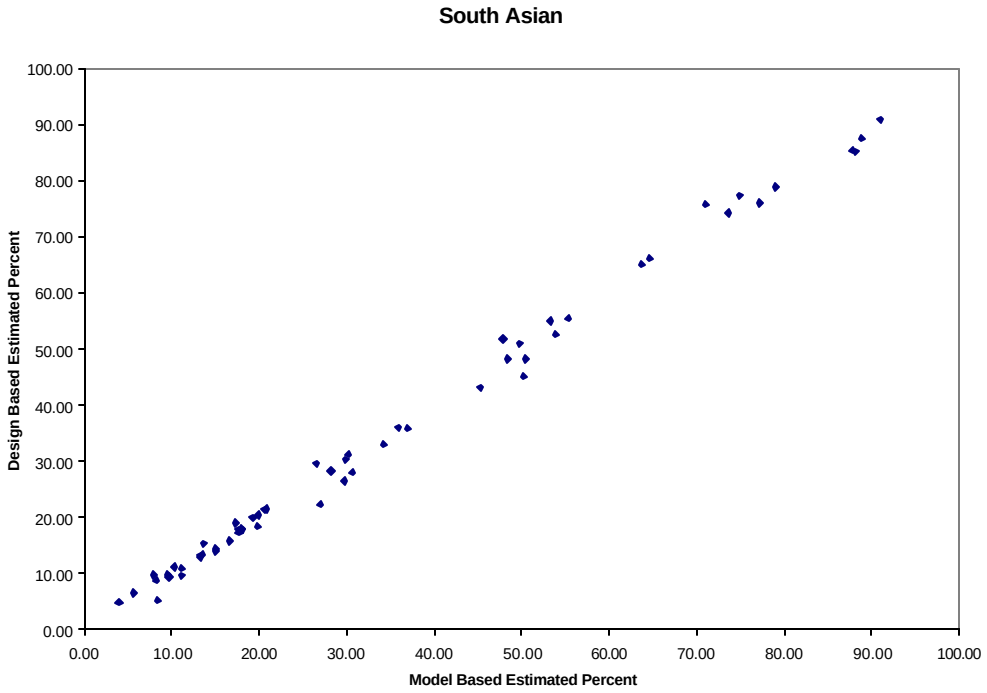

Korean
Design Based Standard Error of a Percentage
Model Based Standard Error of a Percentage
0.000
1.000
2.000
3.000
4.000
5.000
6.000
7.000

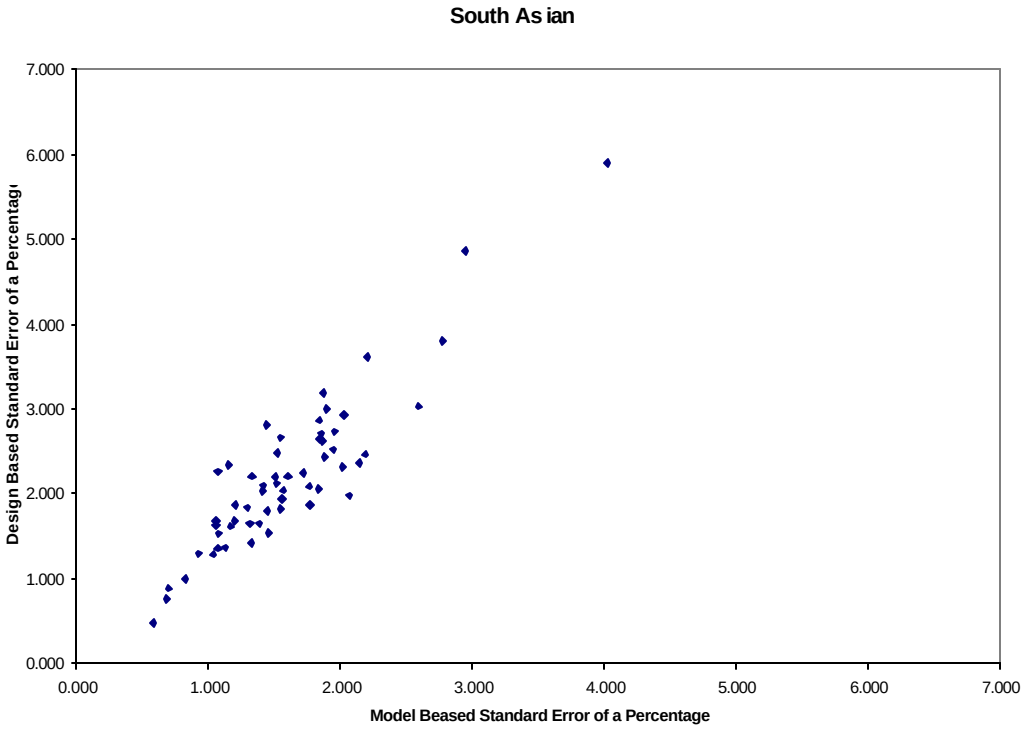

Vietnamese
Design Based Estimated Percent
Model Based Estimated Percent
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00

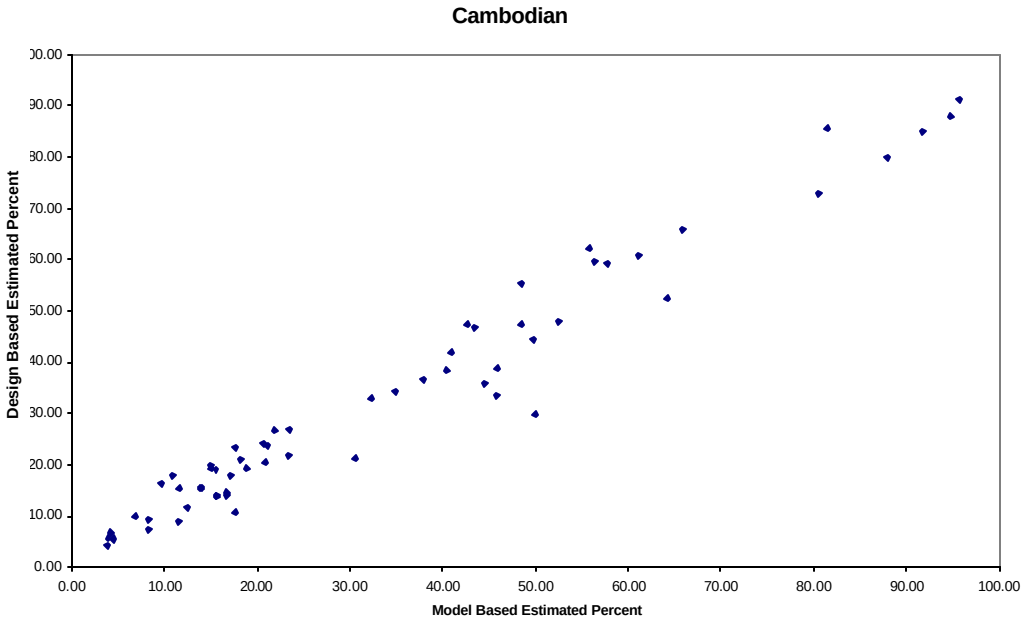

Vietnamese
Design Based Standard Error of a Percentage
Model Based Standard Error of a Percentage
0.000
1.000
2.000
3.000
4.000
5.000
6.000
7.000
8.000

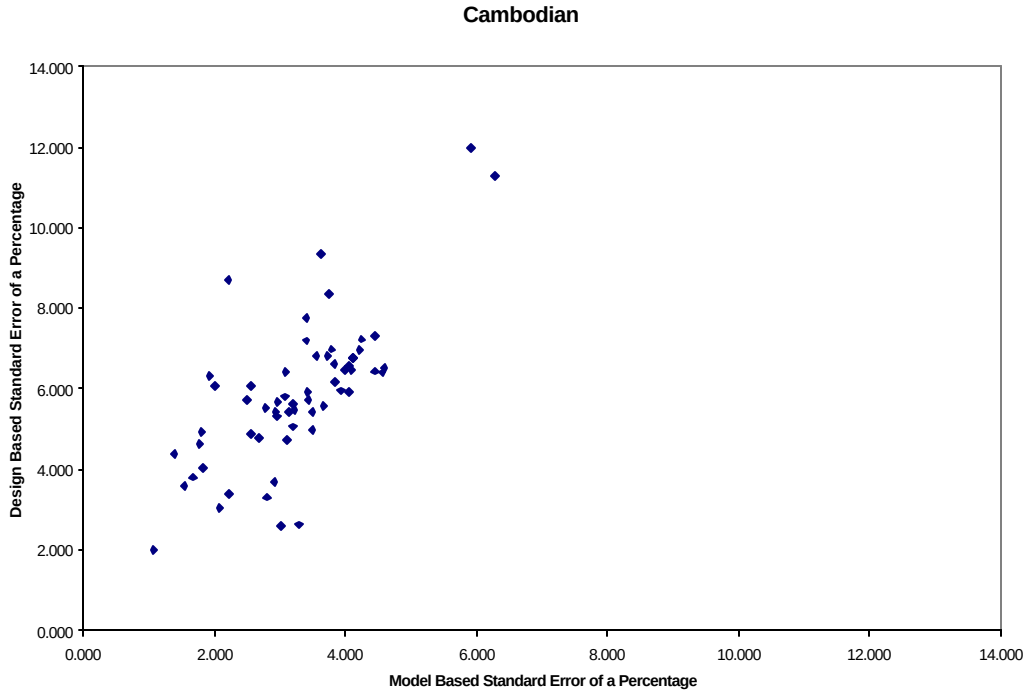

American Indian/Alaskan Native
Design Based Estimated Percen
Model Based Estimated Percent
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00


American Indian/Alaskan Native
Design Based Standard Error of a Percentage
Model Based Standard Error of a Percentage
0.000
1.000
2.000
3.000
4.000
5.000
6.000
7.000
8.000

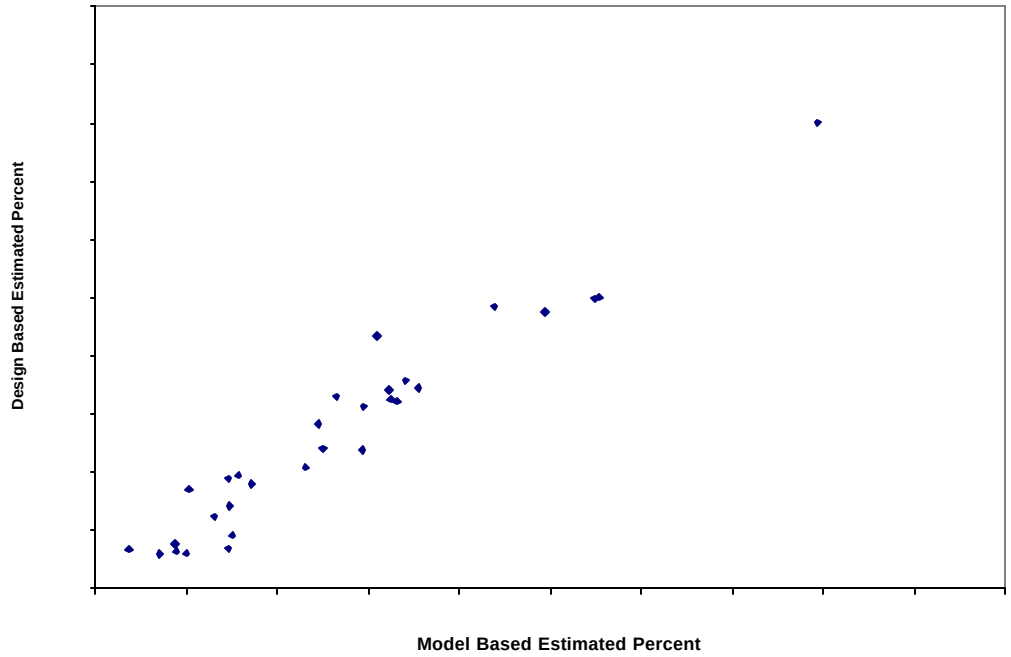

South Asian
Design Based Estimated Percent
Model Based Estimated Percent
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00


South As ian
Design Based Standard Error of a Percentag
Model Beased Standard Error of a Percentage
0.000
1.000
2.000
3.000
4.000
5.000
6.000
7.000

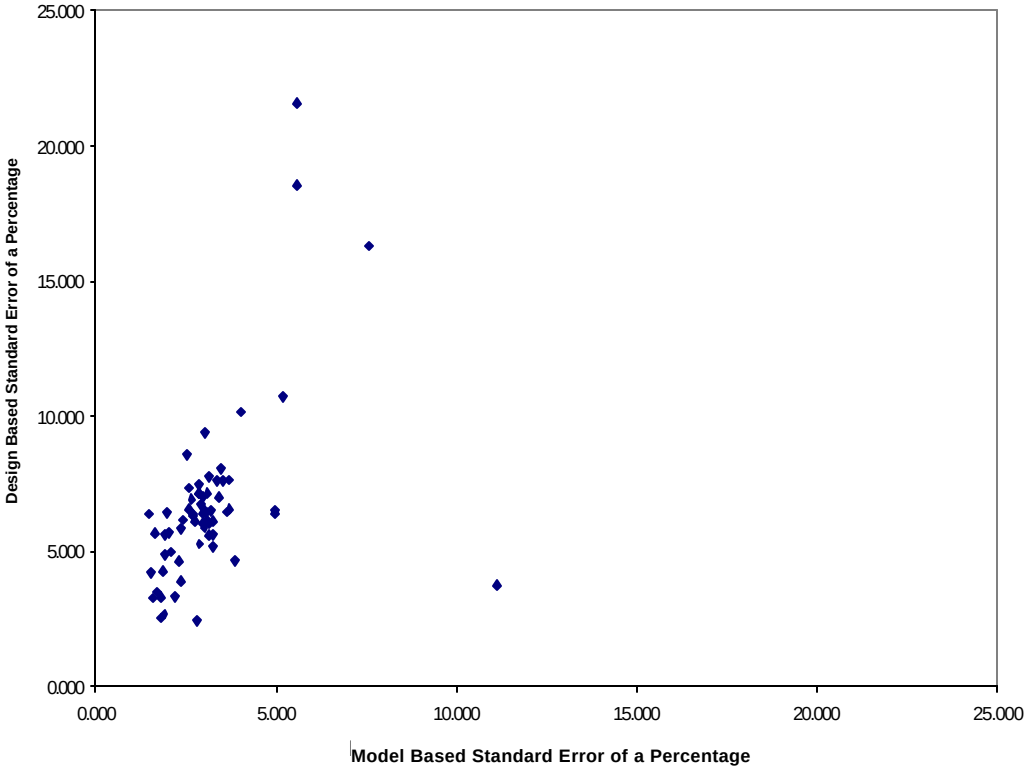

Cambodian
Design Based Estimated Percent
Model Based Estimated Percent
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00


Cambodian
Design Based Standard Error of a Percentage
Model Based Standard Error of a Percentage
0.000
2.000
4.000
6.000
8.000
10.000
12.000
14.000

Shasta Latino



Shasta Latino



10. VARIANCE ESTIMATION

This chapter describes the methods and results of computing sampling errors for CHIS 2001. The data from the RDD and the geographic supplemental samples are covered. 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 2001. The remainder of the chapter describes the methodology for producing estimates of sampling variability from the survey. Section 10.2 is a general review of the two main methods of computing sampling errors or variances of estimates from surveys with complex survey designs like CHIS 2001. Section 10.3 describes the replication method of variance estimation used in the survey. Section 10.4 shows how analysts can compute sampling errors for CHIS 2001 data using commercially available software.

10.1 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

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

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 2001, 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 resulted 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 2001 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 2001 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 2001 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 DEFFs 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 10-1 to 10-3 gives the average, median, maximum, and minimum DEFFs 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 10-1 shows the DEFFs and DEFTs for 39 items selected from the adult interview by the county or stratum reported in the adult interview. Tables 10-2 and 10-3 present the corresponding DEFFs and DEFTs for 23 items from the child interviews and 26 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 B.

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 and O'Muircheartaigh (1980) for a discussion of the use of the DEFT. The main reason for presenting the DEFTs here is because it dampens some of noise associated with the DEFFs. The maximum and minimum values of the DEFFs 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 10-1, the average DEFF for Berkeley is 2.37, while the maximum is 3.98 and the minimum is 0.94. This value is unusually large given the other values in the table. The average DEFT for Berkeley is 1.52, 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 2001. 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, but this had a relatively minor effect on the design effects.

- **Within-Household Subsampling.** Only one adult and one child or adolescent was sampled in each household. This subsampling contributed to the differential weights at the person level because households with more persons were subsampled at different rates.

- **Weighting Adjustments.** Differential weights were applied to reduce nonresponse bias and to make the estimates consistent with known population totals from the 2000 census. 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 10-1 shows the average DEFTs for estimates of adult items are between 1.07 to 1.21 in most counties. This implies that for most counties the standard error of the estimate is about 7 to 21 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 (and Berkeley). The reason for the larger DEFT for Alameda county is that the city of Berkeley was sampled at a much higher rate than the remainder of the county. When the data from the city and the remainder of the county are combined, the cases have very different sampling rates and this increases the DEFT. Similarly, the DEFT for the city of Berkeley is also larger than most other areas because the adults reporting to live in Berkeley were sampled from telephone numbers classified in both Berkeley and the remainder of Alameda county at the time of sampling. The average DEFT for the state estimate is 1.28. This is slightly larger than the county-level DEFTs as expected because counties were not sampled proportional to their population. See Report 1: Sample Design for more details on the sampling.

Table 10-1. Average DEFF and DEFT for estimates from the adult interview*

County/strata	Design effect (DEFF)				DEFT
	Average	Median	Maximum	Minimum	Average
State Total	1.68	1.64	2.94	0.38	1.28
Los Angeles	1.39	1.35	2.29	0.49	1.17
Long Beach	1.38	1.41	2.01	0.48	1.16
Pasadena	1.44	1.50	2.20	0.48	1.19
Remainder of Los Angeles	1.35	1.32	2.17	0.50	1.15
San Diego	1.26	1.35	1.89	0.46	1.11
Orange	1.33	1.31	2.02	0.48	1.14
Santa Clara	1.37	1.36	2.15	0.46	1.16
San Bernardino	1.23	1.22	2.05	0.40	1.10
Riverside	1.26	1.30	1.67	0.37	1.11
Alameda	1.99	1.98	3.63	0.68	1.39
Berkeley	2.37	2.33	3.98	0.94	1.52
Remainder of Alameda	1.35	1.35	2.27	0.45	1.15
Sacramento	1.23	1.25	1.63	0.59	1.10
Contra Costa	1.33	1.33	2.91	0.43	1.14
Fresno	1.49	1.49	2.55	0.44	1.21
San Francisco	1.43	1.50	1.82	0.52	1.19
Ventura	1.38	1.34	2.32	0.42	1.16
San Mateo	1.38	1.38	2.87	0.45	1.16
Kern	1.22	1.20	2.03	0.45	1.10
San Joaquin	1.34	1.28	2.40	0.41	1.15
Sonoma	1.26	1.22	2.09	0.44	1.11
Stanislaus	1.34	1.33	3.85	0.56	1.14
Santa Barbara	1.41	1.38	2.49	0.48	1.17
Solano	1.34	1.30	1.94	0.50	1.15
Tulare	1.46	1.45	2.66	0.61	1.19
Santa Cruz	1.25	1.23	1.95	0.62	1.11
Marin	1.38	1.39	2.80	0.66	1.16
San Luis Obispo	1.23	1.30	2.02	0.38	1.10
Placer	1.17	1.16	2.02	0.37	1.07
Merced	1.34	1.38	2.04	0.55	1.15
Butte	1.32	1.35	3.20	0.42	1.13
Shasta	1.34	1.28	2.43	0.51	1.15
Yolo	1.39	1.40	2.76	0.52	1.16
El Dorado	1.38	1.38	1.94	0.37	1.17
Imperial	1.30	1.27	2.00	0.38	1.13
Napa	1.33	1.31	2.29	0.44	1.14
Kings	1.41	1.35	2.57	0.51	1.17
Madera	1.40	1.38	2.10	0.46	1.17
Monterey, San Benito	1.28	1.38	2.18	0.46	1.12
Del Norte, Humboldt	1.28	1.31	2.13	0.41	1.12
Lassen, Modoc, Siskiyou, Trinity	1.32	1.29	2.86	0.52	1.14
Lake, Mendocino	1.26	1.28	1.93	0.42	1.11

Table 10-1. Average DEFF and DEFT for estimates from the adult interview* (continued)

County/strata	Design effect (DEFF)				DEFT
	Average	Median	Maximum	Minimum	Average
Colusa, Glen, Tehama	1.34	1.31	3.63	0.31	1.14
Sutter, Yuba	1.24	1.21	1.74	0.55	1.11
Plumas, Nevada, Sierra	1.34	1.28	3.21	0.45	1.14
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.22	1.26	1.95	0.35	1.10

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.

Table 10-2 shows the average DEFT for estimates from the child interview in each county. The average DEFTs for estimates from the child interview are larger than those for the adult interview. This result is expected because the subsampling at the person level for children is typically more variable than it is for adults (the number of children per household is more variable than the number of adults per household). The average DEFT at the state level is 1.33. The average DEFTs for the counties are between 1.04 and 1.27; that is, the standard errors of these estimates are between 4 and 27 percent greater than expected from a simple random sample. The only exceptions are again the city of Berkeley and Alameda County for the same reason as noted above for the adult interview items.

Table 10-2. Average DEFF and DEFT for estimates from the child interview*

County/strata	Design effect (DEFF)				DEFT
	Average	Median	Maximum	Minimum	Average
State Total	1.81	1.90	2.31	0.44	1.33
Los Angeles	1.52	1.53	2.31	0.43	1.22
Long Beach	1.28	1.28	2.06	0.27	1.12
Pasadena	1.58	1.47	3.76	0.41	1.22
Remainder of Los Angeles	1.45	1.44	2.28	0.44	1.19
San Diego	1.47	1.36	2.82	0.37	1.19
Orange	1.42	1.37	3.42	0.56	1.17
Santa Clara	1.47	1.42	3.22	0.22	1.18
San Bernardino	1.64	1.65	2.83	0.32	1.26
Riverside	1.45	1.47	2.52	0.21	1.18
Alameda	1.93	2.02	3.25	0.50	1.37
Berkeley	2.24	2.27	4.26	0.08	1.45
Remainder of Alameda	1.53	1.60	2.49	0.40	1.22
Sacramento	1.28	1.27	2.33	0.18	1.11
Contra Costa	1.30	1.36	1.96	0.10	1.12
Fresno	1.68	1.59	3.30	0.21	1.27

Table 10-2. Average DEFF and DEFT for estimates from the child interview* (continued)

County/strata	Design effect (DEFF)				DEFT
	Average	Median	Maximum	Minimum	Average
San Francisco	1.44	1.40	2.35	0.67	1.18
Ventura	1.42	1.46	2.31	0.18	1.17
San Mateo	1.32	1.40	1.89	0.11	1.12
Kern	1.36	1.29	2.32	0.15	1.14
San Joaquin	1.33	1.40	1.84	0.15	1.13
Sonoma	1.25	1.36	2.11	0.01	1.08
Stanislaus	1.23	1.30	1.88	0.11	1.09
Santa Barbara	1.35	1.34	1.94	0.09	1.14
Solano	1.35	1.23	2.36	0.34	1.14
Tulare	1.36	1.37	2.01	0.34	1.15
Santa Cruz	1.35	1.30	1.87	0.08	1.14
Marin	1.12	1.10	2.33	0.16	1.03
San Luis Obispo	1.03	1.12	1.50	0.14	1.00
Placer	1.20	1.19	1.80	0.42	1.09
Merced	1.54	1.57	3.00	0.06	1.21
Butte	1.30	1.40	1.83	0.31	1.12
Shasta	1.33	1.37	1.89	0.01	1.13
Yolo	1.30	1.29	2.35	0.01	1.10
El Dorado	1.35	1.36	2.32	0.08	1.13
Imperial	1.36	1.31	4.03	0.41	1.14
Napa	1.46	1.31	3.23	0.56	1.19
Kings	1.43	1.40	2.18	0.25	1.18
Madera	1.25	1.29	1.74	0.22	1.10
Monterey, San Benito	1.58	1.52	3.47	0.25	1.23
Del Norte, Humboldt	1.61	1.59	2.88	0.10	1.24
Lassen, Modoc, Siskiyou, Trinity	1.34	1.32	2.41	0.15	1.13
Lake, Mendocino	1.33	1.37	1.93	0.02	1.12
Colusa, Glen, Tehama	1.20	1.26	1.66	0.22	1.08
Sutter, Yuba	1.38	1.40	2.48	0.07	1.15
Plumas, Nevada, Sierra	1.15	1.18	1.61	0.04	1.04
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.35	1.43	1.80	0.52	1.15

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.

Table 10-3 shows 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 1.37. For most of the strata, the average DEFTs are between 1.01 and 1.30.

Table 10-3. Average DEFF and DEFT for estimates from the adolescent interview*

County/strata	Design effect (DEFF)				DEFT
	Average	Median	Maximum	Minimum	Average
State Total	1.90	1.99	2.51	1.03	1.37
Los Angeles	1.60	1.58	2.27	0.83	1.26
Long Beach	1.63	1.65	2.47	1.07	1.27
Pasadena	1.05	1.01	1.69	0.32	1.01
Remainder of Los Angeles	1.60	1.58	2.35	0.85	1.26
San Diego	1.36	1.32	1.81	0.71	1.16
Orange	1.31	1.35	1.82	0.30	1.13
Santa Clara	1.31	1.30	2.13	0.77	1.13
San Bernardino	1.54	1.50	2.62	0.94	1.23
Riverside	1.42	1.31	2.75	0.82	1.18
Alameda	1.94	1.96	2.66	0.72	1.38
Berkeley	1.66	1.42	3.91	0.39	1.23
Remainder of Alameda	1.59	1.62	2.18	0.58	1.25
Sacramento	1.50	1.37	2.37	0.88	1.22
Contra Costa	1.16	1.16	1.83	0.51	1.07
Fresno	1.55	1.54	2.67	0.86	1.23
San Francisco	1.52	1.52	3.04	0.32	1.20
Ventura	1.36	1.42	2.04	0.47	1.15
San Mateo	1.30	1.34	1.90	0.27	1.12
Kern	1.48	1.42	2.24	0.90	1.21
San Joaquin	1.42	1.49	2.13	0.77	1.18
Sonoma	1.54	1.44	2.84	0.35	1.21
Stanislaus	1.27	1.20	2.17	0.75	1.12
Santa Barbara	1.40	1.37	2.10	0.44	1.17
Solano	1.47	1.45	2.34	0.79	1.20
Tulare	1.77	1.65	3.95	0.66	1.30
Santa Cruz	1.47	1.44	2.07	0.99	1.21
Marin	1.17	1.22	1.43	0.77	1.08
San Luis Obispo	1.26	1.29	1.77	0.63	1.11
Placer	1.89	1.60	3.31	0.79	1.35
Merced	1.73	1.79	2.55	0.50	1.29
Butte	1.15	1.08	2.18	0.31	1.06
Shasta	1.48	1.51	2.53	0.78	1.21
Yolo	1.28	1.26	1.92	0.63	1.12
El Dorado	1.50	1.54	2.17	0.68	1.21
Imperial	1.31	1.38	1.79	0.63	1.14
Napa	1.46	1.41	2.17	0.77	1.20
Kings	1.62	1.51	3.63	0.61	1.25
Madera	1.40	1.30	2.61	0.82	1.17
Monterey, San Benito	1.41	1.31	2.09	0.74	1.18
Del Norte, Humboldt	1.34	1.34	2.63	0.51	1.15
Lassen, Modoc, Siskiyou, Trinity	1.47	1.51	2.21	0.91	1.21
Lake, Mendocino	1.65	1.76	2.48	0.39	1.26

Table 10-3. Average DEFF and DEFT for estimates from the adolescent interview* (continued)

County/strata	Design effect (DEFF)				DEFT
	Average	Median	Maximum	Minimum	Average
Colusa, Glen, Tehama	1.42	1.25	2.76	0.69	1.17
Sutter, Yuba	1.38	1.39	2.22	0.50	1.16
Plumas, Nevada, Sierra	1.18	1.13	1.67	0.77	1.08
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.69	1.50	3.49	0.93	1.28

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*Includes the RDD sample and San Francisco and Santa Barbara supplemental samples.

10.2 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 take the following form

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

where

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

In the next section, the specific form of equation (1) used in CHIS 2001 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.

10.3 Design of Replicates

In CHIS 2001, a form of the jackknife replication method (JK2[1]) was selected as the preferred method 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.

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

The two major reasons for choosing replication as the primary method to estimate variances for CHIS 2001 were 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 choosing replication is probably more important. Both the nonresponse and raking types of adjustments made in developing the CHIS 2001 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, and 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.

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 2001, 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 80th pair to stratum 80, the 81st pair to stratum 1, etc.). As a result, each variance stratum had approximately the same number of telephone numbers. The same process was followed for each sampling stratum.

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 are 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).

10.4 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 2001 using some of these programs beginning with the program used to compute the sampling errors in this report, WesVar.

WesVar Version 4.1 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 2001 data, the only requirements are to identify the full and replicate weights that are on the data file and specify that the replication method as JK2. This specification is made when a workbook is opened. All of the standard errors produced will properly account for the sample design and estimation methods because these features are accounted for in the replicate weights.

SUDAAN (Software for the Statistical Analysis of Correlated Data) 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.[2] When the Taylor series approximations are used, SUDAAN does not fully take into

[2] 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.

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

STATA has a family of *svy*- commands to analyze data from sample surveys. 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 DEFFs for proportions can be produced using **svytab.** The command **svymean** can be used to produce the DEFFs 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 strata). The variable TSVAR indicates the variance strata 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 PSU's within the sampling strata.

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

REFERENCES

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* (expected to appear in June edition).

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.

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

Deville, J.C., Särndal, C.E., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

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.*

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 Anderson, D.W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society*, A, 149, 65-82.

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.

Lundström, S. and Särndal, C-E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15, 305-327.

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.

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.

SAS (1999). *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.

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

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

Sudman, S., Sirken, M.G., and Cowan, C.D. (1988). Sampling rare and elusive populations. *Science*, 240, 991-996.

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.

WESVAR (2000). *WesVar ™ 4.0 User's Guide,* Westat, Rockville, MD.

Wolter, K. (1985). *Introduction to Variance Estimation.* Springer-Verlag.

Appendix A

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum

		All strata	Long Beach	Pasadena	Remainder of Los Angeles	San Diego	Orange	Santa Clara	San Bernardino	Riverside
1	Base weight									
	1.1 Sample size	345,136	5,511	6,104	74,343	15,626	18,061	10,879	8,607	7,707
	1.2 Sum of weights	27,636,050	352,000	196,100	7,039,900	2,010,600	2,400,200	1,576,050	1,088,350	1,033,550
2	Adjusting for subsampling									
	2.1 Counts after subsampling									
	a. Listed	38,753	420	240	7,315	2,255	1,868	1,186	1,207	1,153
	b. Nonlisted	45,803	518	288	11,552	2,669	2,723	1,496	1,584	1,378
	c. Nonresidential	28,556	289	174	5,878	1,407	1,543	915	810	699
	d. Mail*	96,454	1,854	2,311	21,534	4,359	4,724	2,874	2,068	1,979
	e. Non-mail*	52,470	951	1,268	11,148	1,844	2,886	1,819	1,059	958
	f. Nonresidential*	42,792	828	1,108	7,822	1,291	2,023	1,213	811	632
	2.2 Sample size	304,828	4,860	5,389	65,249	13,825	15,767	9,503	7,539	6,799
	2.3 Sum of weights after adjustment	27,636,050	352,000	196,100	7,039,900	2,010,600	2,400,200	1,576,050	1,088,350	1,033,550
	2.4 Coefficient of Variation (CV)	52.78	11.08	11.18	11.02	11.16	10.80	11.06	11.00	10.93
	2.5 Observed sampling rates									
	- July 2000 sample									
	a. Listed sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	b. Nonlisted sampling rate	79%	79%	77%	78%	78%	79%	78%	78%	79%
	c. Nonresidential sampling rate	82%	81%	83%	83%	83%	81%	83%	83%	82%
	- February 2001 sample									
	d. Mail sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	e. Non-mail sampling rate	80%	80%	80%	80%	80%	80%	80%	80%	81%
	f. Nonresidential sampling rate	80%	80%	80%	80%	80%	80%	79%	80%	80%
3	Adjusting for unknown residential									
	3.1 Sum of weights by residential status before adjustment	27,636,050	352,000	196,100	7,039,900	2,010,600	2,400,200	1,576,050	1,088,350	1,033,550
	a. Residential	10,977,492	133,709	67,925	2,831,491	861,782	889,289	573,565	474,545	456,579
	b. Nonresidential	13,579,731	182,985	101,925	3,452,902	932,112	1,249,026	771,109	514,946	474,286
	c. Unknown residential status – NA, NM	3,078,828	35,306	26,249	755,507	216,706	261,885	231,376	98,860	102,686

*Counts include the supplemental samples from San Francisco and Santa Barbara

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		All strata	Long Beach	Pasadena	Remainder of Los Angeles	San Diego	Orange	Santa Clara	San Bernardino	Riverside
	3.2 Sum of weights before adjustment	12,022,353	145,909	75,976	3,068,683	932,140	963,986	636,514	507,534	491,159
	a. Ineligible	312	0	0	95	0	133	0	0	0
	b. Residential - respondent	7,064,449	87,096	41,943	1,741,293	557,890	568,450	362,745	323,362	305,132
	c. Residential - nonrespondent	3,864,593	46,612	25,982	1,090,104	303,892	320,706	210,820	151,182	151,447
	d. Unknown residential status – NA, NM	1,014,728	12,201	8,051	237,192	70,358	74,697	62,949	32,989	34,581
	e. Residential – outside area	48,137	0	0	0	0	0	0	0	0
	f. Residential – unknown eligibility	30,134	0	0	0	0	0	0	0	0
	3.3 Sum of weights after adjustments	12,022,353	145,909	75,976	3,068,683	932,140	963,986	636,514	507,534	491,159
	a. Ineligible	312	0	0	95	0	133	0	0	0
	b. Residential - respondent	7,706,826	94,980	46,876	1,885,671	602,919	615,429	402,100	345,731	328,139
	c. Residential - nonrespondent	4,232,479	50,929	29,101	1,182,918	329,221	348,424	234,414	161,803	163,020
	d. Residential - outside strata	82,736	0	0	0	0	0	0	0	0
4	Screener nonresponse adjustment									
	4.1 Sum of weights before adjustment	11,939,305	145,909	75,976	3,068,589	932,140	963,853	636,514	507,534	491,159
	a. Respondents	7,706,826	94,980	46,876	1,885,671	602,919	615,429	402,100	345,731	328,139
	b. Nonrespondents	4,232,479	50,929	29,101	1,182,918	329,221	348,424	234,414	161,803	163,020
	4.2 Sum of weights after adjustment	11,939,305	145,909	75,976	3,068,589	932,140	963,853	636,514	507,534	491,159
5	Multiple telephone adjustment									
	5.1 Sum of weights before adjustment	11,939,305	145,909	75,976	3,068,589	932,140	963,853	636,514	507,534	491,159
	5.2 Sum of weights after adjustment	11,231,985	139,348	70,704	2,886,538	878,825	901,377	587,754	485,217	469,439
	5.3 Adjustment factor	0.94	0.96	0.93	0.94	0.94	0.94	0.92	0.96	0.96
6	Household poststratification									
	6.1 Number of completed screeners	84,051	1,287	1,255	17,126	4,055	3,992	2,351	2,350	2,110
	6.2 Sum of weights before adjustment	11,231,985	139,348	70,704	2,886,538	878,825	901,377	587,754	485,217	469,439
	6.3 Sum of weights after adjustment	11,502,870	141,394	71,844	2,920,536	994,677	935,287	565,863	528,594	506,218
	6.4 Adjustment factor	1.02	1.01	1.02	1.01	1.13	1.04	0.96	1.09	1.08
	6.5 CV	60.04	20.26	23.06	23.53	22.67	24.52	25.93	20.45	20.55
7	Section H nonresponse adjustment									
	7.1 Sum of weights before adjustment	11,502,870	141,394	71,844	2,920,536	994,677	935,287	565,863	528,594	506,218
	a. Completed Section H	7,498,548	89,797	46,585	1,811,192	655,214	587,946	364,209	347,132	334,171
	b. Did not complete Section H	4,004,322	51,596	25,258	1,109,344	339,463	347,341	201,654	181,462	172,047
	7.2 Sum of weights after adjustment	11,502,870	141,394	71,844	2,920,536	994,677	935,287	565,863	528,594	506,218
	7.3 Number of adult completes through Section H	55,653	821	817	10,641	2,676	2,510	1,522	1,551	1,394
	7.4 CV	64.44	25.38	25.36	26.20	23.42	27.09	30.68	25.75	23.48

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		Berkeley	Remainder of Alameda	Sacramento	Contra Costa	Fresno	San Francisco	Ventura	San Mateo	Kern
1	Base weight									
	1.1 Sample size	5,017	7,815	7,084	7,354	7,191	20,329	5,825	6,938	6,021
	1.2 Sum of weights	115,700	1,243,600	1,026,950	849,500	631,650	1,204,550	532,050	666,700	509,100
2	Adjusting for subsampling									
	2.1 Counts after subsampling									
	a. Listed	87	847	974	986	709	749	872	910	760
	b. Nonlisted	100	1,143	1,270	1,201	892	913	1,028	1,097	915
	c. Nonresidential	69	701	711	650	820	648	554	695	757
	d. Mail*	2,150	2,109	1,705	2,032	1,769	6,548	1,391	1,714	1,370
	e. Non-mail*	1,389	1,110	890	862	875	6,216	775	832	651
	f. Nonresidential*	675	929	655	759	1,171	4,235	493	815	798
	2.2 Sample size	4,470	6,839	6,205	6,490	6,236	19,309	5,113	6,063	5,251
	2.3 Sum of weights after adjustment	115,700	1,243,600	1,026,950	849,500	631,650	1,204,550	532,050	666,700	509,100
	2.4 Coefficient of Variation (CV)	10.93	10.99	10.85	11.08	10.80	29.98	10.99	11.08	10.61
	2.5 Observed sampling rates									
	- July 2000 sample									
	a. Listed sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	b. Nonlisted sampling rate	79%	78%	79%	78%	80%	79%	79%	79%	79%
	c. Nonresidential sampling rate	82%	83%	82%	83%	80%	81%	83%	82%	81%
	- February 2001 sample									
	d. Mail sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	e. Non-mail sampling rate	80%	80%	80%	80%	80%	80%	80%	78%	81%
	f. Nonresidential sampling rate	81%	80%	80%	80%	79%	80%	80%	80%	80%
3	Adjusting for unknown residential									
	3.1 Sum of weights by residential status before adjustment	115,700	1,243,600	1,026,950	849,500	631,650	1,204,550	532,050	666,700	509,100
	a. Residential	39,716	476,330	424,837	343,950	229,974	390,493	226,431	240,371	199,008
	b. Nonresidential	62,223	609,643	483,851	402,418	352,137	613,371	249,044	340,910	270,137
	c. Unknown residential status – NA, NM	13,760	157,628	118,262	103,132	49,538	200,685	56,576	85,418	39,955

*Counts include the supplemental samples from San Francisco and Santa Barbara

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		Berkeley	Remainder of Alameda	Sacramento	Contra Costa	Fresno	San Francisco	Ventura	San Mateo	Kern
	3.2 Sum of weights before adjustment	44,629	525,300	461,803	378,610	243,573	474,727	243,614	268,571	211,019
	a. Ineligible	0	0	0	0	0	0	0	0	85
	b. Residential - respondent	27,642	300,315	282,808	218,092	155,824	190,549	144,681	145,731	145,068
	c. Residential - nonrespondent	12,075	176,014	142,029	125,859	74,151	153,153	81,750	94,641	53,856
	d. Unknown residential status – NA, NM	4,913	48,970	36,966	34,659	13,599	54,100	17,183	28,200	12,011
	e. Residential – outside area	0	0	0	0	0	46,791	0	0	0
	f. Residential – unknown eligibility	0	0	0	0	0	30,134	0	0	0
	3.3 Sum of weights after adjustments	44,629	525,300	461,803	378,609	243,573	474,727	243,614	268,571	211,019
	a. Ineligible	0	0	0	0	0	0	0	0	85
	b. Residential - respondent	31,066	330,846	307,263	239,712	165,012	219,034	155,360	162,485	153,775
	c. Residential - nonrespondent	13,563	194,454	154,540	138,898	78,561	174,408	88,254	106,087	57,159
	d. Residential - outside strata	0	0	0	0	0	81,285	0	0	0
4	Screener nonresponse adjustment									
	4.1 Sum of weights before adjustment	44,629	525,300	461,803	378,609	243,573	393,442	243,614	268,571	210,935
	a. Respondents	31,066	330,846	307,263	239,712	165,012	219,034	155,359	162,485	153,775
	b. Nonrespondents	13,563	194,454	154,540	138,898	78,561	174,408	88,254	106,087	57,159
	4.2 Sum of weights after adjustment	44,629	525,300	461,803	378,609	243,573	393,442	243,614	268,571	210,935
5	Multiple telephone adjustment									
	5.1 Sum of weights before adjustment	44,629	525,300	461,803	378,609	243,573	393,442	243,614	268,571	210,935
	5.2 Sum of weights after adjustment	40,816	492,528	441,226	351,845	232,419	359,866	228,941	246,279	200,490
	5.3 Adjustment factor	0.91	0.94	0.96	0.93	0.95	0.91	0.94	0.92	0.95
6	Household poststratification									
	6.1 Number of completed screeners	1,155	1,768	1,810	1,777	1,658	3,195	1,485	1,433	1,597
	6.2 Sum of weights before adjustment	40,816	492,528	441,226	351,845	232,419	359,866	228,941	246,279	200,490
	6.3 Sum of weights after adjustment	40,061	483,305	453,602	344,129	252,940	329,700	243,234	254,103	208,652
	6.4 Adjustment factor	0.98	0.98	1.03	0.98	1.09	0.92	1.06	1.03	1.04
	6.5 CV	23.59	24.00	23.34	26.27	19.86	37.45	24.98	27.48	20.71
7	Section H nonresponse adjustment									
	7.1 Sum of weights before adjustment	40,061	483,305	453,602	344,129	252,940	329,700	243,234	254,103	208,652
	a. Completed Section H	27,819	326,077	309,839	233,211	159,528	208,057	158,468	164,982	142,673
	b. Did not complete Section H	12,242	157,228	143,763	110,918	93,412	121,643	84,766	89,121	65,980
	7.2 Sum of weights after adjustment	40,061	483,305	453,602	344,129	252,940	329,700	243,234	254,103	208,652
	7.3 Number of adult completes through Section H	796	1,195	1,239	1,204	1,048	2,006	976	928	1,098
	7.4 CV	24.69	24.93	28.17	28.69	23.55	39.81	31.13	30.70	23.35

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		San Joaquin	Sonoma	Stanislas	Santa Barbara	Solano	Tulare	Santa Cruz	Marin	San Luis Obispo
1	Base weight									
	1.1 Sample size	5,409	3,965	3,941	5,209	7,961	5,022	4,806	4,981	4,135
	1.2 Sum of weights	371,700	417,200	297,900	328,000	271,850	269,500	251,750	307,100	221,050
2	Adjusting for subsampling									
	2.1 Counts after subsampling									
	a. Listed	808	641	578	622	567	453	622	602	656
	b. Nonlisted	1,024	605	573	628	700	463	571	575	541
	c. Nonresidential	610	322	405	315	286	615	372	384	374
	d. Mail*	1,273	1,070	1,103	1,602	3,262	1,155	1,301	1,440	1,144
	e. Non-mail*	577	530	438	923	1,338	541	720	771	502
	f. Nonresidential*	453	359	384	618	977	1,110	640	625	460
	2.2 Sample size	4,745	3,527	3,481	4,708	7,130	4,337	4,226	4,397	3,677
	2.3 Sum of weights after adjustment	371,700	417,200	297,900	328,000	271,850	269,500	251,750	307,100	221,050
	2.4 Coefficient of Variation (CV)	11.00	10.75	11.55	24.14	11.28	10.50	11.04	11.09	11.01
	2.5 Observed sampling rates									
	- July 2000 sample									
	a. Listed sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	b. Nonlisted sampling rate	78%	80%	77%	79%	79%	79%	80%	77%	78%
	c. Nonresidential sampling rate	83%	81%	84%	81%	83%	81%	81%	84%	84%
	- February 2001 sample									
	d. Mail sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	e. Non-mail sampling rate	80%	81%	78%	80%	80%	80%	79%	81%	80%
	f. Nonresidential sampling rate	80%	81%	80%	81%	80%	80%	80%	80%	82%
3	Adjusting for unknown residential									
	3.1 Sum of weights by residential status before adjustment	371,700	417,200	297,900	327,000	271,850	269,500	251,750	307,100	221,050
	a. Residential	173,975	176,272	140,447	138,554	126,565	97,411	93,755	103,719	92,066
	b. Nonresidential	164,206	193,680	131,744	159,144	114,361	156,058	126,460	162,709	107,805
	c. Unknown residential status – NA, NM	33,519	47,248	25,708	30,303	30,924	16,031	31,536	40,672	21,179

*Counts include the supplemental samples from San Francisco and Santa Barbara

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		San Joaquin	Sonoma	Stanislas	Santa Barbara	Solano	Tulare	Santa Cruz	Marin	San Luis Obispo
	3.2 Sum of weights before adjustment	184,763	198,217	151,379	152,564	141,114	104,055	108,150	122,701	101,787
	a. Ineligible	0	0	0	0	0	0	0	0	0
	b. Residential - respondent	119,611	121,358	99,394	93,179	86,727	70,466	62,304	66,608	62,870
	c. Residential - nonrespondent	54,364	54,915	41,054	44,029	39,838	26,945	31,451	37,111	29,196
	d. Unknown residential status – NA, NM	10,788	21,944	10,932	14,011	14,550	6,644	14,396	18,982	9,721
	e. Residential – outside area	0	0	0	1,346	0	0	0	0	0
	f. Residential – unknown eligibility	0	0	0	0	0	0	0	0	0
	3.3 Sum of weights after adjustments	184,763	198,217	151,379	152,564	141,114	104,055	108,150	122,701	101,787
	a. Ineligible	0	0	0	0	0	0	0	0	0
	b. Residential - respondent	126,998	135,837	106,930	102,383	96,421	75,181	71,629	77,825	69,250
	c. Residential - nonrespondent	57,766	62,380	44,449	48,730	44,694	28,874	36,521	44,876	32,538
	d. Residential - outside strata	0	0	0	1,451	0	0	0	0	0
4	Screener nonresponse adjustment									
	4.1 Sum of weights before adjustment	184,763	198,217	151,379	151,113	141,114	104,055	108,150	122,701	101,787
	a. Respondents	126,997	135,837	106,930	102,383	96,421	75,181	71,629	77,825	69,250
	b. Nonrespondents	57,766	62,380	44,449	48,730	44,694	28,874	36,521	44,876	32,538
	4.2 Sum of weights after adjustment	184,763	198,217	151,379	151,113	141,114	104,055	108,150	122,701	101,787
5	Multiple telephone adjustment									
	5.1 Sum of weights before adjustment	184,763	198,217	151,379	151,113	141,114	104,055	108,150	122,701	101,787
	5.2 Sum of weights after adjustment	176,180	185,190	144,225	142,853	132,614	99,913	98,946	110,657	95,859
	5.3 Adjustment factor	0.95	0.93	0.95	0.95	0.94	0.96	0.91	0.90	0.94
6	Household poststratification									
	6.1 Number of completed screeners	1,609	1,095	1,228	1,452	2,415	1,232	1,123	1,040	1,113
	6.2 Sum of weights before adjustment	176,180	185,190	144,225	142,853	132,614	99,913	98,946	110,657	95,859
	6.3 Sum of weights after adjustment	181,629	172,403	145,146	136,622	130,403	110,385	91,139	100,650	92,739
	6.4 Adjustment factor	1.03	0.93	1.01	0.96	0.98	1.10	0.92	0.91	0.97
	6.5 CV	22.81	29.21	24.25	36.03	26.51	23.52	30.92	40.33	26.68
7	Section H nonresponse adjustment									
	7.1 Sum of weights before adjustment	181,629	172,403	145,146	136,622	130,403	110,385	91,139	100,650	92,739
	a. Completed Section H	119,123	120,678	96,334	95,269	85,973	73,806	64,883	72,082	66,426
	b. Did not complete Section H	62,506	51,725	48,812	41,353	44,430	36,579	26,256	28,568	26,313
	7.2 Sum of weights after adjustment	181,629	172,403	145,146	136,622	130,403	110,385	91,139	100,650	92,739
	7.3 Number of adult completes through Section H	1,062	771	824	1,011	1,596	831	795	750	801
	7.4 CV	26.43	30.81	27.68	36.06	29.95	27.26	30.80	42.00	32.82

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		Placer	Merced	Butte	Shasta	Yolo	El Dorado	Imperial	Napa	Kings
1	Base weight									
	1.1 Sample size	4,280	4,094	3,478	3,766	3,754	4,425	3,947	4,557	4,306
	1.2 Sum of weights	222,850	114,800	151,350	136,050	118,450	122,050	71,600	109,500	66,450
2	Adjusting for subsampling									
	2.1 Counts after subsampling									
	a. Listed	664	517	744	662	646	569	530	613	524
	b. Nonlisted	574	557	525	544	562	489	630	584	599
	c. Nonresidential	334	350	320	367	361	375	386	366	422
	d. Mail*	1,086	1,193	888	941	927	1,282	1,112	1,309	1,155
	e. Non-mail*	731	498	356	485	430	609	453	664	490
	f. Nonresidential*	381	503	273	341	382	588	384	500	583
	2.2 Sample size	3,770	3,618	3,106	3,340	3,308	3,912	3,495	4,036	3,773
	2.3 Sum of weights after adjustment	222,850	114,800	151,350	136,050	118,450	122,050	71,600	109,500	66,450
	2.4 Coefficient of Variation (CV)	11.20	11.06	11.38	10.96	11.34	10.99	11.05	11.10	11.33
	2.5 Observed sampling rates									
	- July 2000 sample									
	a. Listed sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	b. Nonlisted sampling rate	78%	79%	79%	79%	80%	80%	78%	78%	77%
	c. Nonresidential sampling rate	83%	82%	82%	82%	80%	80%	84%	84%	84%
	- February 2001 sample									
	d. Mail sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	e. Non-mail sampling rate	79%	80%	81%	80%	80%	80%	81%	81%	79%
	f. Nonresidential sampling rate	81%	80%	78%	82%	78%	80%	81%	80%	80%
3	Adjusting for unknown residential									
	3.1 Sum of weights by residential status before adjustment	222,850	114,800	151,350	136,050	118,450	122,050	71,600	109,500	66,450
	a. Residential	93,574	53,183	75,082	62,673	53,413	49,416	32,517	44,401	29,561
	b. Nonresidential	103,556	52,574	60,599	59,784	53,177	57,635	31,848	51,035	31,502
	c. Unknown residential status – NA, NM	25,720	9,043	15,669	13,593	11,860	14,999	7,235	14,064	5,387

*Counts include the supplemental samples from San Francisco and Santa Barbara

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		Placer	Merced	Butte	Shasta	Yolo	El Dorado	Imperial	Napa	Kings
	3.2 Sum of weights before adjustment	104,607	57,505	82,265	68,485	58,584	56,248	35,611	50,443	31,883
	a. Ineligible	0	0	0	0	0	0	0	0	0
	b. Residential - respondent	62,934	38,004	55,294	44,924	38,743	32,431	23,773	29,634	20,809
	c. Residential - nonrespondent	30,641	15,179	19,788	17,749	14,670	16,985	8,744	14,767	8,752
	d. Unknown residential status – NA, NM	11,033	4,322	7,184	5,812	5,171	6,832	3,094	6,042	2,322
	e. Residential – outside area	0	0	0	0	0	0	0	0	0
	f. Residential – unknown eligibility	0	0	0	0	0	0	0	0	0
	3.3 Sum of weights after adjustments	104,607	57,505	82,265	68,485	58,584	56,248	35,611	50,443	31,883
	a. Ineligible	0	0	0	0	0	0	0	0	0
	b. Residential - respondent	70,212	41,103	60,491	49,002	42,297	36,742	25,990	33,409	22,439
	c. Residential - nonrespondent	34,396	16,402	21,775	19,483	16,287	19,506	9,621	17,034	9,443
	d. Residential - outside strata	0	0	0	0	0	0	0	0	0
4	Screener nonresponse adjustment									
	4.1 Sum of weights before adjustment	104,607	57,505	82,265	68,485	58,584	56,248	35,611	50,443	31,883
	a. Respondents	70,212	41,103	60,491	49,002	42,297	36,742	25,990	33,409	22,439
	b. Nonrespondents	34,396	16,402	21,775	19,483	16,287	19,506	9,621	17,034	9,443
	4.2 Sum of weights after adjustment	104,607	57,505	82,265	68,485	58,584	56,248	35,611	50,443	31,883
5	Multiple telephone adjustment									
	5.1 Sum of weights before adjustment	104,607	57,505	82,265	68,485	58,584	56,248	35,611	50,443	31,883
	5.2 Sum of weights after adjustment	99,750	55,213	77,793	65,631	54,985	53,843	34,244	47,378	30,487
	5.3 Adjustment factor	0.95	0.96	0.95	0.96	0.94	0.96	0.96	0.94	0.96
6	Household poststratification									
	6.1 Number of completed screeners	1,128	1,265	1,198	1,170	1,163	1,113	1,225	1,174	1,249
	6.2 Sum of weights before adjustment	99,750	55,213	77,793	65,631	54,985	53,843	34,244	47,378	30,487
	6.3 Sum of weights after adjustment	93,382	63,815	79,566	63,426	59,375	58,939	39,384	45,402	34,418
	6.4 Adjustment factor	0.94	1.16	1.02	0.97	1.08	1.09	1.15	0.96	1.13
	6.5 CV	25.08	20.72	23.53	23.68	28.42	29.34	22.78	31.79	22.73
7	Section H nonresponse adjustment									
	7.1 Sum of weights before adjustment	93,382	63,815	79,566	63,426	59,375	58,939	39,384	45,402	34,418
	a. Completed Section H	65,038	41,683	54,618	44,217	42,888	41,319	25,615	30,932	23,135
	b. Did not complete Section H	28,344	22,132	24,948	19,209	16,487	17,621	13,769	14,470	11,283
	7.2 Sum of weights after adjustment	93,382	63,815	79,566	63,426	59,375	58,939	39,384	45,402	34,418
	7.3 Number of adult completes through Section H	787	834	825	827	834	784	801	809	846
	7.4 CV	26.01	24.31	26.25	31.24	29.35	32.04	25.01	32.99	28.96

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	Colusa, Glen, Tehama	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1	Base weight									
	1.1 Sample size	4,091	5,433	5,034	5,256	4,723	4,107	4,357	4,535	5,152
	1.2 Sum of weights	76,150	376,450	145,650	97,950	96,700	97,050	101,750	116,100	202,550
2	Adjusting for subsampling									
	2.1 Counts after subsampling									
	a. Listed	559	414	554	579	603	577	573	720	618
	b. Nonlisted	579	555	394	383	495	440	517	483	446
	c. Nonresidential	414	434	603	594	463	418	464	382	500
	d. Mail*	954	1,492	1,186	1,209	1,233	1,058	1,129	1,183	1,276
	e. Non-mail*	612	832	615	644	583	561	529	715	790
	f. Nonresidential*	454	1,012	1,011	1,151	770	558	627	520	870
	2.2 Sample size	3,572	4,739	4,363	4,560	4,147	3,612	3,839	4,003	4,500
	2.3 Sum of weights after adjustment	76,150	376,450	145,650	97,950	96,700	97,050	101,750	116,100	202,550
	2.4 Coefficient of Variation (CV)	10.92	10.51	10.89	10.64	10.89	11.07	10.65	11.19	10.85
	2.5 Observed sampling rates									
	- July 2000 sample									
	a. Listed sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	b. Nonlisted sampling rate	79%	79%	78%	79%	79%	78%	81%	80%	78%
	c. Nonresidential sampling rate	82%	81%	81%	80%	81%	82%	79%	80%	82%
	- February 2001 sample									
	d. Mail sample rate	100%	100%	100%	100%	100%	100%	100%	100%	100%
	e. Non-mail sampling rate	80%	80%	79%	80%	80%	81%	82%	80%	79%
	f. Nonresidential sampling rate	80%	81%	79%	80%	80%	79%	80%	79%	81%
3	Adjusting for unknown residential									
	3.1 Sum of weights by residential status before adjustment	76,150	376,450	145,650	97,950	96,700	97,050	101,750	116,100	202,550
	a. Residential	33,456	129,111	50,980	31,709	36,863	41,364	42,460	44,320	70,650
	b. Nonresidential	37,228	207,507	83,119	58,013	49,722	47,119	50,792	56,253	109,079
	c. Unknown residential status – NA, NM	5,466	39,833	11,551	8,228	10,115	8,568	8,498	15,527	22,821

*Counts include the supplemental samples from San Francisco and Santa Barbara

Table A-1. Household weighting for the combined RDD sample and geographic supplemental samples by stratum (continued)

		Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	Colusa, Glen, Tehama	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
	3.2 Sum of weights before adjustment	35,638	146,276	56,182	35,304	41,523	44,857	45,946	51,475	81,042
	a. Ineligible	0	0	0	0	0	0	0	0	0
	b. Residential - respondent	24,130	88,688	36,702	23,405	25,220	30,886	30,353	30,506	46,878
	c. Residential - nonrespondent	9,326	40,422	14,278	8,304	11,644	10,478	12,107	13,814	23,773
	d. Unknown residential status – NA, NM	2,183	17,166	5,202	3,595	4,660	3,494	3,486	7,155	10,391
	e. Residential – outside area	0	0	0	0	0	0	0	0	0
	f. Residential – unknown eligibility	0	0	0	0	0	0	0	0	0
	3.3 Sum of weights after adjustments	35,638	146,276	56,182	35,304	41,523	44,857	45,946	51,475	81,042
	a. Ineligible	0	0	0	0	0	0	0	0	0
	b. Residential - respondent	25,642	100,361	40,425	26,017	28,401	33,489	32,834	35,364	53,762
	c. Residential - nonrespondent	9,997	45,916	15,757	9,286	13,122	11,369	13,112	16,111	27,280
	d. Residential - outside strata	0	0	0	0	0	0	0	0	0
4	Screener nonresponse adjustment									
	4.1 Sum of weights before adjustment	35,638	146,276	56,182	35,304	41,523	44,857	45,946	51,475	81,042
	a. Respondents	25,642	100,361	40,425	26,017	28,401	33,489	32,834	35,364	53,762
	b. Nonrespondents	9,997	45,916	15,757	9,287	13,122	11,369	13,112	16,111	27,280
	4.2 Sum of weights after adjustment	35,638	146,276	56,182	35,304	41,523	44,857	45,946	51,475	81,042
5	Multiple telephone adjustment									
	5.1 Sum of weights before adjustment	35,638	146,276	56,182	35,304	41,523	44,857	45,946	51,475	81,042
	5.2 Sum of weights after adjustment	34,412	137,544	53,358	34,200	39,366	43,384	43,967	48,467	77,916
	5.3 Adjustment factor	0.97	0.94	0.95	0.97	0.95	0.97	0.96	0.94	0.96
6	Household poststratification									
	6.1 Number of completed screeners	1,207	1,201	1,197	1,191	1,159	1,229	1,228	1,130	1,113
	6.2 Sum of weights before adjustment	34,412	137,544	53,358	34,200	39,366	43,384	43,967	48,467	77,916
	6.3 Sum of weights after adjustment	36,155	137,121	60,408	37,552	57,240	36,282	47,568	47,414	70,168
	6.4 Adjustment factor	1.05	1.00	1.13	1.10	1.45*	0.84*	1.08	0.98	0.90
	6.5 CV	23.04	26.25	23.87	22.95	22.24	20.55	21.03	24.27	21.85
7	Section H nonresponse adjustment									
	7.1 Sum of weights before adjustment	36,155	137,121	60,408	37,552	57,240	36,282	47,568	47,414	70,168
	a. Completed Section H	24,633	90,398	43,314	26,349	40,180	24,835	31,964	34,191	51,768
	b. Did not complete Section H	11,522	46,723	17,094	11,203	17,060	11,447	15,604	13,223	18,400
	7.2 Sum of weights after adjustment	36,155	137,121	60,408	37,552	57,240	36,282	47,568	47,414	70,168
	7.3 Number of adult completes through Section H	827	792	863	847	813	841	824	815	821
	7.4 CV	27.19	25.66	27.37	27.43	21.87	20.69	24.34	27.60	22.88

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

*The poststratification adjustment factors for the strata containing Lake County and Colusa County would have been 1.16 and 0.90 respectively, if the correct base weights for these two strata were used. The use of the factors in this table corrected for the error in the base weights. See Chapter 3, Table 3-1 for further details.

Table A-2. Extended interview weighting for adult interview by stratum (combined RDD sample and geographic supplemental samples)

		All strata	Long Beach	Pasadena	Remainder of Los Angeles	San Diego	Orange	Santa Clara	San Bernardino	Riverside
1	Adult initial weights									
	1.1 Number of sampled adults	84,051	1,287	1,255	17,126	4,055	3,992	2,351	2,350	2,110
	1.2 Sum of weights	22,901,507	262,668	136,152	5,903,281	1,950,693	1,918,643	1,167,808	1,098,764	1,032,970
	1.3 CV	81.78	52.54	51.92	54.51	51.17	51.40	53.80	51.73	52.05
2	Nonresponse adjustment									
	2.1 Number of completed adult interviews	55,430	819	814	10,582	2,666	2,495	1,514	1,547	1,386
	2.2 Sum of weights before adjustment	22,901,507	262,668	136,152	5,903,281	1,950,693	1,918,643	1,167,808	1,098,764	1,032,970
	a. Eligible respondents	14,330,675	159,596	85,116	3,519,932	1,231,353	1,156,453	709,106	699,123	668,995
	b. Ineligibles	18,099	0	0	6,664	1,405	0	600	3,935	461
	c. Nonrespondents	8,552,734	103,073	51,035	2,376,685	717,935	762,190	458,102	395,706	363,514
	2.3 Sum of weights after adjustment	22,901,508	262,668	136,152	5,903,281	1,950,693	1,918,643	1,167,808	1,098,764	1,032,970
	2.4 Mean adjustment factor	1.60	1.65	1.60	1.68	1.58	1.66	1.65	1.57	1.54
3	Trimming adjustment**									
	3.1 Number of trimmed records	9	0	0	0	0	1	0	0	0
	3.2 Sum of weights before trimming adjustment	22,870,309	338,093	123,333	5,828,564	1,952,317	1,877,229	1,154,716	1,096,389	1,033,076
	3.3 Sum of weights after trimming adjustment	22,865,817	338,093	123,333	5,828,564	1,952,317	1,876,539	1,154,716	1,096,389	1,033,076
4	Raking nonresponse adjustment									
	4.1 Number of completed adult interviews	55,430	913	671	10,612	2,672	2,454	1,510	1,554	1,391
	4.2 Sum of weights after adjustment	23,848,253	317,193	99,595	6,271,792	1,996,400	2,038,769	1,238,203	1,114,747	1,044,523
	4.3 Mean adjustment factor	1.04	1.21	0.73	1.06	1.03	1.06	1.06	1.02	1.01
	4.4 CV	96.64	78.28	80.69	66.33	64.90	66.43	67.60	65.09	61.46
	4.5 Mean weight	430.24	347.42	148.43	591.01	747.16	830.79	820.00	717.34	750.92

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

Table A-2. Extended interview weighting for adult interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Berkeley	Remainder of Alameda	Sacramento	Contra Costa	Fresno	San Francisco	Ventura	San Mateo	Kern
1	Adult initial weights									
	1.1 Number of sampled adults	1,155	1,768	1,810	1,777	1,658	3,195	1,485	1,433	1,597
	1.2 Sum of weights	70,444	946,521	840,339	690,023	507,183	611,767	499,505	509,130	403,705
	1.3 CV	49.32	53.29	47.57	51.93	50.69	62.77	52.86	54.97	47.04
2	Nonresponse adjustment									
	2.1 Number of completed adult interviews	794	1,191	1,238	1,199	1,041	1,995	971	925	1,096
	2.2 Sum of weights before adjustment	70,444	946,521	840,339	690,023	507,183	611,767	499,505	509,130	403,705
	a. Eligible respondents	46,640	612,836	549,911	449,550	303,132	361,561	315,880	308,171	266,751
	b. Ineligibles	0	0	0	0	0	498	0	174	0
	c. Nonrespondents	23,805	333,685	290,428	240,473	204,051	249,708	183,625	200,786	136,953
	2.3 Sum of weights after adjustment	70,444	946,521	840,339	690,023	507,183	611,767	499,505	509,130	403,705
	2.4 Mean adjustment factor	1.51	1.54	1.53	1.53	1.67	1.69	1.58	1.65	1.51
3	Trimming adjustment**									
	3.1 Number of trimmed records	1	0	0	0	0	0	0	1	0
	3.2 Sum of weights before trimming adjustment	113,532	909,231	825,284	683,076	509,168	599,673	521,881	534,556	402,099
	3.3 Sum of weights after trimming adjustment	112,792	909,231	825,284	683,076	509,168	599,673	521,881	533,940	402,099
4	Raking nonresponse adjustment									
	4.1 Number of completed adult interviews	809	1,165	1,231	1,214	1,053	1,955	1,015	972	1,093
	4.2 Sum of weights after adjustment	82,505	980,325	862,200	686,521	526,605	644,954	526,821	535,285	421,106
	4.3 Mean adjustment factor	1.17	1.04	1.03	1.00	1.04	1.06	1.06	1.05	1.04
	4.4 CV	146.23	66.99	61.75	66.19	75.01	73.58	71.33	72.60	63.42
	4.5 Mean weight	101.98	841.48	700.41	565.50	500.10	329.90	519.04	550.71	385.28

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

Table A-2. Extended interview weighting for adult interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano	Tulare	Santa Cruz	Marin	San Luis Obispo
1	Adult initial weights									
	1.1 Number of sampled adults	1,609	1,095	1,228	1,452	2,415	1,232	1,123	1,040	1,113
	1.2 Sum of weights	365,014	319,636	293,611	272,269	263,453	227,734	181,262	184,926	175,599
	1.3 CV	51.01	51.56	51.70	61.99	52.48	55.06	57.15	57.21	49.60
2	Nonresponse adjustment									
	2.1 Number of completed adult interviews	1,052	771	819	1,004	1,587	827	793	750	799
	2.2 Sum of weights before adjustment	365,014	319,636	293,611	272,269	263,453	227,734	181,262	184,926	175,599
	a. Eligible respondents	231,499	215,073	186,493	179,231	167,940	146,045	124,765	130,081	121,907
	b. Ineligibles	0	821	0	226	158	0	0	355	0
	c. Nonrespondents	133,516	103,742	107,118	92,812	95,355	81,690	56,497	54,490	53,692
	2.3 Sum of weights after adjustment	365,014	319,636	293,611	272,269	263,453	227,734	181,262	184,926	175,599
	2.4 Mean adjustment factor	1.58	1.49	1.57	1.52	1.57	1.56	1.45	1.42	1.44
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	0	0	0	0	0	0	2
	3.2 Sum of weights before trimming adjustment	366,895	318,813	281,766	272,843	258,405	228,304	183,580	184,947	177,638
	3.3 Sum of weights after trimming adjustment	366,895	318,813	281,766	272,843	258,405	228,304	183,580	184,947	176,580
4	Raking nonresponse adjustment									
	4.1 Number of completed adult interviews	1,058	776	794	1,001	1,553	826	791	755	807
	4.2 Sum of weights after adjustment	372,075	336,506	301,247	283,900	267,150	238,351	186,125	186,111	178,255
	4.3 Mean adjustment factor	1.02	1.06	1.03	1.04	1.02	1.05	1.03	1.01	1.02
	4.4 CV	66.41	66.19	69.62	75.60	65.71	78.42	68.44	65.90	61.69
	4.5 Mean weight	351.68	433.64	379.40	283.62	172.02	288.56	235.30	246.50	220.89

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

Table A-2. Extended interview weighting for adult interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Placer	Merced	Butte	Shasta	Yolo	El Dorado	Imperial	Napa	Kings
1	Adult initial weights									
	1.1 Number of sampled adults	1,128	1,265	1,198	1,170	1,163	1,113	1,225	1,174	1,249
	1.2 Sum of weights	179,364	132,857	148,480	117,866	118,410	111,798	85,423	87,481	68,837
	1.3 CV	47.58	55.15	48.15	45.65	51.43	50.00	55.82	61.10	47.09
2	Nonresponse adjustment									
	2.1 Number of completed adult interviews	784	832	825	826	834	780	798	806	843
	2.2 Sum of weights before adjustment	179,364	132,857	148,480	117,866	118,410	111,798	85,423	87,481	68,837
	a. Eligible respondents	121,913	83,931	99,896	80,638	82,816	75,003	53,775	57,486	45,414
	b. Ineligibles	200	204	0	0	0	265	90	0	107
	c. Nonrespondents	57,252	48,722	48,584	37,227	35,594	36,531	31,558	29,996	23,316
	2.3 Sum of weights after adjustment	179,364	132,857	148,480	117,866	118,410	111,798	85,423	87,481	68,837
	2.4 Mean adjustment factor	1.47	1.58	1.49	1.46	1.43	1.49	1.59	1.52	1.52
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	0	0	2	0	0	0	0
	3.2 Sum of weights before trimming adjustment	174,074	142,227	150,305	117,954	127,075	124,318	84,965	91,807	69,717
	3.3 Sum of weights after trimming adjustment	174,074	142,227	150,305	117,954	126,256	124,318	84,965	91,807	69,717
4	Raking nonresponse adjustment									
	4.1 Number of completed adult interviews	764	849	835	827	844	807	794	833	837
	4.2 Sum of weights after adjustment	179,895	135,474	148,804	117,566	119,101	114,580	86,796	89,281	71,883
	4.3 Mean adjustment factor	1.01	1.02	1.00	1.00	1.01	1.03	1.02	1.02	1.05
	4.4 CV	58.14	71.40	66.03	62.67	79.22	80.32	72.83	76.39	80.29
	4.5 Mean weight	235.46	159.57	178.21	142.16	141.11	141.98	109.32	107.18	85.88

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

Table A-2. Extended interview weighting for adult interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	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	1,207	1,201	1,197	1,191	1,159	1,229	1,228	1,130	1,113
	1.2 Sum of weights	72,907	277,127	111,077	70,691	106,586	69,215	92,642	86,823	130,823
	1.3 CV	47.15	50.46	47.43	49.24	50.71	48.75	50.55	46.63	46.46
2	Nonresponse adjustment									
	2.1 Number of completed adult interviews	824	790	861	846	813	839	822	814	818
	2.2 Sum of weights before adjustment	72,907	277,127	111,077	70,691	106,586	69,215	92,642	86,823	130,823
	a. Eligible respondents	48,780	173,679	76,787	48,391	73,303	45,647	60,128	61,174	94,777
	b. Ineligibles	0	1,062	222	202	0	158	0	293	0
	c. Nonrespondents	24,127	102,386	34,069	22,098	33,283	23,410	32,513	25,357	36,046
	2.3 Sum of weights after adjustment	72,907	277,127	111,077	70,691	106,586	69,215	92,642	86,823	130,823
	2.4 Mean adjustment factor	1.49	1.60	1.45	1.46	1.45	1.52	1.54	1.42	1.38
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	2	0	0	0	0	0	0
	3.2 Sum of weights before trimming adjustment	72,612	274,009	110,831	70,416	106,042	69,191	90,534	88,660	130,166
	3.3 Sum of weights after trimming adjustment	72,612	274,009	110,264	70,416	106,042	69,191	90,534	88,660	130,166
4	Raking nonresponse adjustment									
	4.1 Number of completed adult interviews	820	794	855	841	808	839	801	824	813
	4.2 Sum of weights after adjustment	78,748	303,354	110,080	67,104	105,599	70,242	94,953	88,566	128,964
	4.3 Mean adjustment factor	1.08	1.10	0.99	0.95	0.99	1.02	1.03	1.03	0.99
	4.4 CV	71.17	73.76	67.58	62.23	63.76	65.99	62.68	60.82	58.28
	4.5 Mean weight	96.03	382.06	128.75	79.79	130.69	83.72	118.54	107.48	158.63

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

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Table A-3. Extended interview weighting for child interview by stratum (combined RDD sample and geographic supplemental samples)

		All strata	Long Beach	Pasadena	Remainder of Los Angeles	San Diego	Orange	Santa Clara	San Bernardino	Riverside
1	Child initial weights									
	1.1 Number of sampled children	14,338	234	201	2,939	671	716	392	491	416
	1.2 Sum of weights	6,195,867	85,585	35,628	1,676,661	494,326	538,598	281,062	351,579	330,792
	1.3 CV	101.40	72.36	70.65	70.79	77.98	93.79	69.62	79.15	70.07
2	Nonresponse adjustment									
	2.1 Number of completed child interviews	12,593	202	168	2,454	584	619	353	442	376
	2.2 Sum of weights before adjustment	6,195,867	85,585	35,628	1,676,661	494,326	538,598	281,062	351,579	330,792
	a. Eligible respondents	5,383,397	74,501	28,890	1,398,047	436,789	458,987	256,503	319,328	300,226
	b. Ineligibles	19,494	160	0	7,503	709	0	879	1,135	0
	c. Nonrespondents	792,975	10,925	6,738	271,110	56,829	79,612	23,681	31,116	30,566
	2.3 Sum of weights after adjustment	6,195,866	85,585	35,628	1,676,661	494,326	538,598	281,062	351,579	330,792
	2.4 Mean adjustment factor	1.15	1.15	1.23	1.20	1.13	1.17	1.10	1.10	1.10
3	Trimming adjustment**									
	3.1 Number of trimmed records	24	1	0	1	2	3	0	2	0
	3.2 Sum of weights before trimming adjustment	6,173,292	111,894	27,067	1,655,550	494,888	520,070	280,622	350,115	331,616
	3.3 Sum of weights after trimming adjustment	6,142,598	111,243	27,067	1,654,109	490,246	509,508	280,622	346,950	331,616
4	Raking nonresponse adjustment									
	4.1 Number of completed child interviews	12,593	231	129	2,460	585	604	354	440	378
	4.2 Sum of weights after adjustment	6,251,997	94,646	21,786	1,719,340	492,815	531,691	285,698	371,444	315,681
	4.3 Mean adjustment factor	1.01	1.11	0.61	1.03	1.00	0.99	1.02	1.06	0.95
	4.4 CV	105.46	82.80	83.07	78.65	75.89	73.73	76.71	90.46	79.38
	4.5 Mean weight	496.47	409.72	168.88	698.92	842.42	880.28	807.06	844.19	835.14

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

Table A-3. Extended interview weighting for child interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Berkeley	Remainder of Alameda	Sacramento	Contra Costa	Fresno	San Francisco	Ventura	San Mateo	Kern
1	Child initial weights									
	1.1 Number of sampled children	95	296	344	288	308	308	261	179	366
	1.2 Sum of weights	7,361	243,133	218,876	160,172	176,703	73,255	142,925	110,764	141,217
	1.3 CV	60.26	69.78	60.11	73.93	62.24	67.36	72.23	76.02	59.35
2	Nonresponse adjustment									
	2.1 Number of completed child interviews	90	265	302	263	270	276	225	154	323
	2.2 Sum of weights before adjustment	7,361	243,133	218,876	160,172	176,703	73,255	142,925	110,764	141,217
	a. Eligible respondents	6,842	214,386	185,558	141,814	156,411	65,552	120,336	94,111	125,784
	b. Ineligibles	0	396	389	0	947	128	1,236	0	1,037
	c. Nonrespondents	519	28,352	32,930	18,358	19,345	7,576	21,353	16,653	14,396
	2.3 Sum of weights after adjustment	7,361	243,133	218,876	160,172	176,703	73,255	142,925	110,764	141,217
	2.4 Mean adjustment factor	1.08	1.13	1.18	1.13	1.13	1.12	1.19	1.18	1.12
3	Trimming adjustment**									
	3.1 Number of trimmed records	1	0	1	0	1	0	1	1	0
	3.2 Sum of weights before trimming adjustment	16,528	237,363	212,028	156,492	174,689	72,271	151,322	111,082	139,845
	3.3 Sum of weights after trimming adjustment	16,059	237,363	211,378	156,492	174,160	72,271	150,653	109,673	139,845
4	Raking nonresponse adjustment									
	4.1 Number of completed child interviews	97	265	295	258	270	272	242	157	322
	4.2 Sum of weights after adjustment	9,694	232,911	226,121	168,436	170,982	75,445	143,754	110,373	141,065
	4.3 Mean adjustment factor	1.32	0.96	1.04	1.05	0.97	1.03	1.02	1.00	1.01
	4.4 CV	153.70	73.06	72.55	69.66	87.33	86.20	73.82	82.13	63.50
	4.5 Mean weight	99.94	878.91	766.51	652.85	633.26	277.37	594.03	703.01	438.09

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

Table A-3. Extended interview weighting for child interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano	Tulare	Santa Cruz	Marin	San Luis Obispo
1	Child initial weights									
	1.1 Number of sampled children	314	170	234	248	467	244	198	150	165
	1.2 Sum of weights	109,044	66,044	95,525	64,253	73,266	73,582	39,913	33,497	35,235
	1.3 CV	64.18	61.98	70.47	77.03	66.18	63.48	76.62	90.64	61.91
2	Nonresponse adjustment									
	2.1 Number of completed child interviews	282	161	198	223	403	223	175	133	152
	2.2 Sum of weights before adjustment	109,044	66,044	95,525	64,253	73,266	73,582	39,913	33,497	35,235
	a. Eligible respondents	97,640	62,383	81,077	57,093	63,313	66,671	35,148	28,917	32,675
	b. Ineligibles	618	0	544	0	332	677	0	342	0
	c. Nonrespondents	10,787	3,661	13,904	7,159	9,621	6,234	4,764	4,238	2,559
	2.3 Sum of weights after adjustment	109,044	66,044	95,525	64,253	73,266	73,582	39,913	33,497	35,235
	2.4 Mean adjustment factor	1.12	1.06	1.18	1.13	1.16	1.10	1.14	1.16	1.08
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	1	1	0	0	1	1	0
	3.2 Sum of weights before trimming adjustment	108,489	66,044	89,608	64,591	72,542	74,004	39,583	33,227	34,814
	3.3 Sum of weights after trimming adjustment	108,489	66,044	88,291	63,718	72,542	74,004	39,093	32,891	34,814
4	Raking nonresponse adjustment									
	4.1 Number of completed child interviews	282	161	191	223	400	226	173	134	152
	4.2 Sum of weights after adjustment	114,654	72,089	91,693	66,675	74,102	82,489	39,665	33,344	33,467
	4.3 Mean adjustment factor	1.06	1.09	0.97	1.04	1.02	1.13	0.99	1.01	0.95
	4.4 CV	70.84	62.41	66.85	74.71	75.00	65.46	78.27	55.81	58.01
	4.5 Mean weight	406.57	447.76	480.07	298.99	185.25	365.00	229.28	248.83	220.18

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

Table A-3. Extended interview weighting for child interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Placer	Merced	Butte	Shasta	Yolo	El Dorado	Imperial	Napa	Kings
1	Child initial weights									
	1.1 Number of sampled children	198	262	186	184	209	180	270	182	306
	1.2 Sum of weights	44,796	48,314	32,129	27,199	29,238	25,326	25,383	21,856	25,349
	1.3 CV	58.01	68.86	57.89	70.26	58.72	61.49	69.84	148.88	63.94
2	Nonresponse adjustment									
	2.1 Number of completed child interviews	179	228	169	163	198	167	226	164	276
	2.2 Sum of weights before adjustment	44,796	48,314	32,129	27,199	29,238	25,326	25,383	21,856	25,349
	a. Eligible respondents	40,254	41,173	28,876	23,445	27,728	23,545	20,597	17,449	22,658
	b. Ineligibles	227	627	86	0	71	0	177	54	71
	c. Nonrespondents	4,314	6,514	3,167	3,754	1,439	1,781	4,610	4,353	2,620
	2.3 Sum of weights after adjustment	44,796	48,314	32,129	27,199	29,238	25,326	25,383	21,856	25,349
	2.4 Mean adjustment factor	1.11	1.17	1.11	1.16	1.05	1.08	1.23	1.25	1.12
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	1	0	0	1	1	0	0	0
	3.2 Sum of weights before trimming adjustment	43,671	53,401	32,098	27,321	33,538	29,282	25,165	21,961	25,518
	3.3 Sum of weights after trimming adjustment	43,671	52,257	32,098	27,321	32,597	28,717	25,165	21,961	25,518
4	Raking nonresponse adjustment									
	4.1 Number of completed child interviews	177	236	170	165	203	176	226	166	275
	4.2 Sum of weights after adjustment	42,706	48,027	30,779	26,315	27,918	25,504	28,954	19,403	25,498
	4.3 Mean adjustment factor	0.96	1.01	0.96	0.97	0.96	1.01	1.15	0.89	1.01
	4.4 CV	66.81	90.23	72.97	75.75	72.46	77.45	73.72	69.71	74.32
	4.5 Mean weight	241.28	203.51	181.05	159.49	137.53	144.91	128.11	116.89	92.72

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

Table A-3. Extended interview weighting for child interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	Colusa, Glen, Tehama	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1	Child initial weights									
	1.1 Number of sampled children	222	235	187	159	153	215	197	160	138
	1.2 Sum of weights	23,742	83,445	23,450	15,834	24,862	18,754	26,525	17,413	23,255
	1.3 CV	73.34	68.49	101.80	64.39	71.49	60.76	62.35	74.47	63.44
2	Nonresponse adjustment									
	2.1 Number of completed child interviews	185	210	171	151	140	196	179	149	126
	2.2 Sum of weights before adjustment	23,742	83,445	23,450	15,834	24,862	18,754	26,525	17,413	23,255
	a. Eligible respondents	20,078	72,365	21,504	15,209	21,751	16,879	23,869	15,502	21,535
	b. Ineligibles	279	0	295	0	104	131	130	57	156
	c. Nonrespondents	3,385	11,080	1,651	624	3,008	1,745	2,526	1,854	1,564
	2.3 Sum of weights after adjustment	23,742	83,445	23,450	15,834	24,862	18,754	26,525	17,413	23,255
	2.4 Mean adjustment factor	1.18	1.15	1.09	1.04	1.14	1.11	1.11	1.12	1.08
3	Trimming adjustment**									
	3.1 Number of trimmed records	1	0	2	0	0	0	0	0	0
	3.2 Sum of weights before trimming adjustment	23,155	83,290	22,551	16,290	24,751	18,602	25,664	17,292	23,397
	3.3 Sum of weights after trimming adjustment	23,076	83,290	21,792	16,290	24,751	18,602	25,664	17,292	23,397
4	Raking nonresponse adjustment									
	4.1 Number of completed child interviews	183	210	168	152	140	196	174	148	127
	4.2 Sum of weights after adjustment	23,971	88,616	22,530	14,090	22,342	18,632	27,357	16,038	23,259
	4.3 Mean adjustment factor	1.02	1.06	0.97	0.89	0.90	1.00	1.04	0.92	1.01
	4.4 CV	87.35	81.26	98.53	72.70	68.32	63.57	83.39	55.85	66.79
	4.5 Mean weight	130.99	421.98	134.11	92.70	159.58	95.06	157.22	108.36	183.14

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

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Table A-4. Extended interview weighting for adolescent interview by stratum (combined RDD sample and geographic supplemental samples)

		All strata	Long Beach	Pasadena	Remainder of Los Angeles	San Diego	Orange	Santa Clara	San Bernardino	Riverside
1	Adolescent initial weights									
	1.1 Number of sampled adolescents	9,245	110	95	1,747	443	391	227	334	249
	1.2 Sum of weights	2,967,702	30,218	11,914	770,184	243,005	219,679	131,962	196,538	158,379
	1.3 CV	96.13	57.57	59.55	69.02	54.53	67.98	59.15	79.74	88.09
2	Nonresponse adjustment									
	2.1 Number of completed adolescent interviews	5,801	62	54	1,007	271	217	138	210	157
	2.2 Sum of weights before adjustment	2,967,702	30,218	11,914	770,184	243,005	219,679	131,962	196,538	158,379
	a. Eligible respondents	1,798,536	16,489	6,910	445,289	150,103	111,260	74,661	132,244	98,379
	b. Ineligibles	47,034	0	0	12,544	3,065	3,920	1,982	3,450	4,744
	c. Nonrespondents	1,122,131	13,730	5,003	312,351	89,837	104,499	55,318	60,843	55,256
	2.3 Number of completed adolescent interviews	2,967,702	30,218	11,914	770,184	243,005	219,679	131,962	196,538	158,379
	2.4 Mean adjustment factor	1.65	1.83	1.72	1.73	1.62	1.97	1.77	1.49	1.61
3	Trimming adjustment**									
	3.1 Number of trimmed records	20	1	0	1	0	0	1	2	1
	3.2 Sum of weights before trimming adjustment	2,892,026	39,345	9,191	745,929	235,735	206,738	128,990	192,716	150,328
	3.3 Sum of weights after trimming adjustment	2,867,828	37,448	9,191	743,876	235,735	206,738	128,337	186,636	143,721
4	Raking nonresponse adjustment									
	4.1 Number of completed adolescent interviews	5,801	65	36	1,020	270	212	140	211	157
	4.2 Sum of weights after adjustment	2,951,644	39,502	9,037	771,204	227,604	233,463	128,969	178,212	150,830
	4.3 Mean adjustment factor	1.02	1.31	0.76	1.03	0.96	1.10	1.01	0.93	1.00
	4.4 CV	107.92	87.06	65.11	79.98	60.88	67.79	69.72	76.29	75.90
	4.5 Mean weight	508.82	607.73	251.04	756.08	842.98	1101.24	921.21	844.61	960.70

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

Table A-4. Extended interview weighting for adolescent interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Berkeley	Remainder of Alameda	Sacramento	Contra Costa	Fresno	San Francisco	Ventura	San Mateo	Kern
1	Adolescent initial weights									
	1.1 Number of sampled adolescents	57	172	217	198	221	153	178	124	230
	1.2 Sum of weights	3,384	101,698	107,532	77,652	80,287	31,230	74,227	54,495	67,098
	1.3 CV	47.90	50.06	56.44	56.41	51.58	67.69	65.90	57.40	56.87
2	Nonresponse adjustment									
	2.1 Number of completed adolescent interviews	30	104	143	128	138	82	106	79	149
	2.2 Sum of weights before adjustment	3,384	101,698	107,532	77,652	80,287	31,230	74,227	54,495	67,098
	a. Eligible respondents	1,733	61,649	69,604	48,159	51,202	15,268	41,940	34,995	44,055
	b. Ineligibles	0	654	1,908	1,606	1,918	984	556	298	1,765
	c. Nonrespondents	1,650	39,395	36,020	27,888	27,167	14,978	31,731	19,202	21,278
	2.3 Number of completed adolescent interviews	3,384	101,698	107,532	77,652	80,287	31,230	74,227	54,495	67,098
	2.4 Mean adjustment factor	1.95	1.65	1.54	1.61	1.57	2.05	1.77	1.56	1.52
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	1	0	0	0	0	1	1	1
	3.2 Sum of weights before trimming adjustment	8,927	98,267	97,952	72,158	76,068	28,667	76,340	52,750	63,405
	3.3 Sum of weights after trimming adjustment	8,927	97,410	97,952	72,158	76,068	28,667	75,206	52,110	63,081
4	Raking nonresponse adjustment									
	4.1 Number of completed adolescent interviews	36	105	135	121	136	80	112	77	147
	4.2 Sum of weights after adjustment	4,722	105,850	109,682	82,522	84,153	36,576	69,410	51,052	69,504
	4.3 Mean adjustment factor	1.40	1.05	1.05	1.10	1.09	1.25	0.95	0.95	1.08
	4.4 CV	112.48	82.94	82.59	70.37	84.59	79.24	72.32	71.41	74.26
	4.5 Mean weight	131.15	1008.09	812.46	682.00	618.78	457.21	619.73	663.02	472.82

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

Table A-4. Extended interview weighting for adolescent interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano	Tulare	Santa Cruz	Marin	San Luis Obispo
1	Adolescent initial weights									
	1.1 Number of sampled adolescents	208	135	166	152	283	168	149	114	101
	1.2 Sum of weights	59,808	41,854	50,826	30,183	35,473	39,233	22,374	19,453	16,264
	1.3 CV	58.20	57.93	68.29	77.57	65.12	69.19	71.16	74.49	56.62
2	Nonresponse adjustment									
	2.1 Number of completed adolescent interviews	133	90	95	93	174	107	104	67	67
	2.2 Sum of weights before adjustment	59,808	41,854	50,826	30,183	35,473	39,233	22,374	19,453	16,264
	a. Eligible respondents	38,381	26,280	30,471	18,796	23,147	23,586	15,496	10,271	10,129
	b. Ineligibles	660	437	333	0	781	383	0	1,538	523
	c. Nonrespondents	20,767	15,137	20,022	11,387	11,546	15,264	6,878	7,644	5,611
	2.3 Number of completed adolescent interviews	59,808	41,854	50,826	30,183	35,473	39,233	22,374	19,453	16,264
	2.4 Mean adjustment factor	1.56	1.59	1.67	1.61	1.53	1.66	1.44	1.89	1.61
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	1	1	0	0	1	0	1
	3.2 Sum of weights before trimming adjustment	59,939	41,222	48,131	30,883	34,264	40,181	22,068	17,276	15,845
	3.3 Sum of weights after trimming adjustment	59,939	41,222	47,689	30,689	34,264	40,181	21,105	17,276	15,473
4	Raking nonresponse adjustment									
	4.1 Number of completed adolescent interviews	135	90	92	93	174	110	103	68	68
	4.2 Sum of weights after adjustment	58,098	38,916	46,568	32,155	37,316	41,140	20,784	16,348	19,388
	4.3 Mean adjustment factor	0.99	0.94	0.93	1.07	1.09	1.06	0.93	0.96	1.25
	4.4 CV	75.39	84.90	70.35	91.33	83.56	90.01	69.68	50.32	70.22
	4.5 Mean weight	430.36	432.40	506.17	345.75	214.46	374.00	201.78	240.42	285.12

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

Table A-4. Extended interview weighting for adolescent interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Placer	Merced	Butte	Shasta	Yolo	El Dorado	Imperial	Napa	Kings
1	Adolescent initial weights									
	1.1 Number of sampled adolescents	127	168	112	131	135	133	224	131	210
	1.2 Sum of weights	21,571	23,783	15,253	14,848	15,043	15,896	17,647	12,431	12,274
	1.3 CV	48.52	75.35	49.48	58.97	59.48	54.55	55.11	146.98	65.44
2	Nonresponse adjustment									
	2.1 Number of completed adolescent interviews	88	115	71	87	91	96	154	85	150
	2.2 Sum of weights before adjustment	21,571	23,783	15,253	14,848	15,043	15,896	17,647	12,431	12,274
	a. Eligible respondents	15,174	15,486	9,566	9,359	10,166	11,740	12,435	7,264	8,684
	b. Ineligibles	0	0	0	541	460	75	222	140	47
	c. Nonrespondents	6,397	8,297	5,688	4,949	4,417	4,081	4,990	5,026	3,543
	2.3 Number of completed adolescent interviews	21,571	23,783	15,253	14,848	15,043	15,896	17,647	12,431	12,274
	2.4 Mean adjustment factor	1.42	1.54	1.59	1.59	1.48	1.35	1.42	1.71	1.41
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	0	0	0	3	0	0	0
	3.2 Sum of weights before trimming adjustment	20,430	25,727	15,253	14,042	16,381	22,083	17,334	12,193	13,331
	3.3 Sum of weights after trimming adjustment	20,430	25,727	15,253	14,042	16,381	20,583	17,334	12,193	13,331
4	Raking nonresponse adjustment									
	4.1 Number of completed adolescent interviews	84	117	71	87	94	106	154	85	151
	4.2 Sum of weights after adjustment	22,911	24,198	17,744	16,017	14,126	15,164	15,567	10,362	11,951
	4.3 Mean adjustment factor	1.06	1.02	1.16	1.14	0.98	0.96	0.90	0.85	0.98
	4.4 CV	70.99	103.33	54.68	77.56	74.73	88.73	74.73	82.23	91.41
	4.5 Mean weight	272.75	206.82	249.92	184.10	150.27	143.05	101.09	121.90	79.15

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

Table A-4. Extended interview weighting for adolescent interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	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	157	150	156	120	120	158	131	119	141
	1.2 Sum of weights	12,046	39,517	15,083	9,263	15,050	10,721	13,562	9,449	19,315
	1.3 CV	58.98	54.48	76.30	58.92	77.06	66.03	61.46	52.14	62.98
2	Nonresponse adjustment									
	2.1 Number of completed adolescent	104	100	109	82	82	108	81	92	101
	2.2 Sum of weights before adjustment	12,046	39,517	15,083	9,263	15,050	10,721	13,562	9,449	19,315
	a. Eligible respondents	8,432	25,644	9,667	6,430	10,118	7,367	8,476	7,406	14,626
	b. Ineligibles	42	334	59	102	73	373	437	0	78
	c. Nonrespondents	3,572	13,539	5,357	2,731	4,859	2,981	4,649	2,044	4,611
	2.3 Number of completed adolescent interviews	12,046	39,517	15,083	9,263	15,050	10,721	13,562	9,449	19,315
	2.4 Mean adjustment factor	1.43	1.54	1.56	1.44	1.49	1.46	1.60	1.28	1.32
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	1	1	0	0	0	1	0
	3.2 Sum of weights before trimming adjustment	11,984	39,325	14,990	9,120	14,947	10,205	12,853	9,449	19,064
	3.3 Sum of weights after trimming adjustment	11,984	39,325	14,783	8,992	14,947	10,205	12,853	9,300	19,064
4	Raking nonresponse adjustment									
	4.1 Number of completed adolescent interviews	104	101	109	82	82	108	81	92	100
	4.2 Sum of weights after adjustment	12,290	41,543	13,509	9,150	13,403	10,582	14,122	10,715	15,284
	4.3 Mean adjustment factor	1.03	1.07	0.90	1.00	0.90	1.04	1.10	1.13	0.80
	4.4 CV	79.08	67.18	76.91	71.48	80.10	78.50	88.64	53.08	72.85
	4.5 Mean weight	118.17	411.32	123.94	111.58	163.45	97.98	174.35	116.47	152.84

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

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Table A-5. Extended interview weighting for adolescent insurance interview by stratum (combined RDD sample and geographic supplemental samples)

		All strata	Long Beach	Pasadena	Remainder of Los Angeles	San Diego	Orange	Santa Clara	San Bernardino	Riverside
1	Adolescent insurance initial weights									
	1.1 Number of sampled adolescents	9,245	110	95	1,747	443	391	227	334	249
	1.2 Sum of weights	2,967,702	30,218	11,914	770,184	243,005	219,679	131,962	196,538	158,379
	1.3 CV	96.13	57.57	59.55	69.02	54.53	67.98	59.15	79.74	88.09
2	Nonresponse adjustment									
	2.1 Number of completed insurance interviews	7,796	85	75	1,376	381	307	193	278	208
	2.2 Sum of weights before adjustment	2,957,802	30,218	11,914	770,184	243,005	219,679	131,962	196,538	158,379
	a. Eligible respondents	2,452,633	23,739	9,403	600,813	207,476	170,717	111,813	167,816	130,874
	b. Ineligibles	39,734	0	0	12,544	3,065	3,920	1,982	3,450	4,744
	c. Nonrespondents	465,435	6,480	2,510	156,827	32,464	45,042	18,167	25,271	22,761
	2.3 Number of weights after adjustment	2,967,702	30,218	11,914	770,184	243,005	219,679	131,962	196,538	158,379
	2.4 Mean adjustment factor	1.21	1.27	1.27	1.28	1.17	1.29	1.18	1.17	1.21
3	Trimming adjustment**									
	3.1 Number of trimmed records	17	1	0	1	0	0	0	2	1
	3.2 Sum of weights before trimming adjustment	2,911,863	40,542	9,337	747,395	237,680	210,808	130,122	193,357	152,736
	3.3 Sum of weights after trimming adjustment	2,893,856	39,150	9,337	745,889	237,680	210,808	130,122	188,862	147,815
4	Raking nonresponse adjustment									
	4.1 Number of completed insurance interviews	7,796	92	51	1,387	380	302	195	278	209
	4.2 Sum of weights after adjustment	2,951,644	39,502	9,037	771,204	227,604	233,463	128,969	178,212	150,830
	4.3 Mean adjustment factor	1.01	1.31	0.76	1.02	0.95	1.09	1.00	0.93	0.99
	4.4 CV	106.92	75.05	70.36	77.92	62.17	76.18	80.97	78.34	74.46
	4.5 Mean weight	378.61	429.37	177.20	556.02	598.96	773.06	661.38	641.05	721.67

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

Table A-5. Extended interview weighting for adolescent insurance interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Berkeley	Remainder of Alameda	Sacramento	Contra Costa	Fresno	San Francisco	Ventura	San Mateo	Kern
1	Adolescent insurance initial weights									
	1.1 Number of sampled adolescents	57	172	217	198	221	153	178	124	230
	1.2 Sum of weights	3,384	101,698	107,532	77,652	80,287	31,230	74,227	54,495	67,098
	1.3 CV	47.90	50.06	56.44	56.41	51.58	67.69	65.90	57.40	56.87
2	Nonresponse adjustment									
	2.1 Number of completed insurance interviews	49	146	186	178	183	127	147	105	196
	2.2 Sum of weights before adjustment	2,784	101,098	107,532	77,652	80,287	30,330	73,727	54,295	67,098
	a. Eligible respondents	2,721	87,419	90,499	66,861	66,044	25,283	61,946	46,984	56,516
	b. Ineligibles	0	54	1,908	1,606	1,918	84	56	98	1,765
	c. Nonrespondents	62	13,624	15,125	9,185	12,325	4,963	11,724	7,212	8,817
	2.3 Number of weights after adjustment	3,384	101,698	107,532	77,652	80,287	31,230	74,227	54,495	67,098
	2.4 Mean adjustment factor	1.02	1.16	1.19	1.16	1.22	1.20	1.19	1.16	1.19
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	2	1	0	0	0	1	0	0
	3.2 Sum of weights before trimming adjustment	9,596	98,134	99,332	73,337	77,020	29,343	76,316	53,379	64,146
	3.3 Sum of weights after trimming adjustment	9,596	97,202	98,229	73,337	77,020	29,343	75,598	53,379	64,146
4	Raking nonresponse adjustment									
	4.1 Number of completed insurance interviews	58	146	177	170	182	123	156	104	194
	4.2 Sum of weights after adjustment	4,722	105,850	109,682	82,522	84,153	36,576	69,410	51,052	69,504
	4.3 Mean adjustment factor	1.40	1.05	1.04	1.09	1.08	1.22	0.95	0.94	1.07
	4.4 CV	125.26	71.17	87.38	69.22	79.85	111.35	79.21	69.71	75.36
	4.5 Mean weight	81.41	725.00	619.67	485.42	462.38	297.37	444.93	490.89	358.27

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

Table A-5. Extended interview weighting for adolescent insurance interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		San Joaquin	Sonoma	Stanislaus	Santa Barbara	Solano	Tulare	Santa Cruz	Marin	San Luis Obispo
1	Adolescent insurance initial weights									
	1.1 Number of sampled adolescents	208	135	166	152	283	168	149	114	101
	1.2 Sum of weights	59,808	41,854	50,826	30,183	35,473	39,233	22,374	19,453	16,264
	1.3 CV	58.20	57.93	68.29	77.57	65.12	69.19	71.16	74.49	56.62
2	Nonresponse adjustment									
	2.1 Number of completed insurance interviews	176	120	140	135	228	146	135	98	92
	2.2 Sum of weights before adjustment	59,208	41,454	50,526	30,183	34,773	38,933	22,374	19,453	15,764
	a. Eligible respondents	51,187	36,835	43,815	26,994	28,731	33,794	20,423	15,352	14,162
	b. Ineligibles	60	37	33	0	81	83	0	1,538	23
	c. Nonrespondents	7,961	4,582	6,678	3,189	5,962	5,055	1,950	2,563	1,579
	2.3 Number of weights after adjustment	59,808	41,854	50,826	30,183	35,473	39,233	22,374	19,453	16,264
	2.4 Mean adjustment factor	1.16	1.13	1.15	1.12	1.21	1.15	1.10	1.27	1.11
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	1	0	0	0	0	1	0	1
	3.2 Sum of weights before trimming adjustment	59,915	41,324	47,696	30,744	34,606	40,024	22,094	17,864	15,745
	3.3 Sum of weights after trimming adjustment	59,915	40,904	47,696	30,744	34,606	40,024	21,358	17,864	15,478
4	Raking nonresponse adjustment									
	4.1 Number of completed insurance interviews	178	120	134	135	228	149	133	99	93
	4.2 Sum of weights after adjustment	58,098	38,916	46,568	32,155	37,316	41,140	20,784	16,348	19,388
	4.3 Mean adjustment factor	0.98	0.94	0.92	1.07	1.08	1.06	0.93	0.92	1.24
	4.4 CV	72.32	70.21	73.27	84.57	85.69	90.75	68.35	51.85	67.79
	4.5 Mean weight	326.39	324.30	347.52	238.18	163.67	276.11	156.27	165.14	208.48

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

Table A-5. Extended interview weighting for adolescent insurance interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Placer	Merced	Butte	Shasta	Yolo	El Dorado	Imperial	Napa	Kings
1	Adolescent insurance initial weights									
	1.1 Number of sampled adolescents	127	168	112	131	135	133	224	131	210
	1.2 Sum of weights	21,571	23,783	15,253	14,848	15,043	15,896	17,647	12,431	12,274
	1.3 CV	48.52	75.35	49.48	58.97	59.48	54.55	55.11	146.98	65.44
2	Nonresponse adjustment									
	2.1 Number of completed insurance interviews	112	144	100	113	120	118	196	110	184
	2.2 Sum of weights before adjustment	21,571	23,783	15,253	14,348	14,643	15,896	17,447	12,331	12,274
	a. Eligible respondents	19,156	19,504	13,647	12,549	13,232	14,090	15,673	9,368	10,978
	b. Ineligibles	0	0	0	41	60	75	22	40	47
	c. Nonrespondents	2,415	4,279	1,607	1,758	1,351	1,731	1,752	2,922	1,249
	2.3 Number of weights after adjustment	21,571	23,783	15,253	14,848	15,043	15,896	17,647	12,431	12,274
	2.4 Mean adjustment factor	1.13	1.22	1.12	1.14	1.11	1.13	1.11	1.32	1.12
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	0	0	0	3	0	0	0
	3.2 Sum of weights before trimming adjustment	20,668	26,471	15,253	14,239	16,773	20,737	17,400	12,243	12,905
	3.3 Sum of weights after trimming adjustment	20,668	26,471	15,253	14,239	16,773	19,559	17,400	12,243	12,905
4	Raking nonresponse adjustment									
	4.1 Number of completed insurance interviews	108	150	100	113	124	128	196	111	184
	4.2 Sum of weights after adjustment	22,911	24,198	17,744	16,017	14,126	15,164	15,567	10,362	11,951
	4.3 Mean adjustment factor	1.06	1.02	1.16	1.13	0.97	0.96	0.90	0.85	0.98
	4.4 CV	68.24	98.72	60.87	81.08	85.94	84.17	68.83	74.91	88.62
	4.5 Mean weight	212.14	161.32	177.44	141.74	113.92	118.47	79.43	93.35	64.95

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

Table A-5. Extended interview weighting for adolescent insurance interview by stratum (combined RDD sample and geographic supplemental samples) (continued)

		Madera	Monterey, San Benito	Del Norte, Humboldt	Lassen, Modoc, Siskiyou, Trinity	Lake, Mendocino	Colusa, Glen, Tehama	Sutter, Yuba	Plumas, Nevada, Sierra	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne
1	Adolescent insurance initial weights									
	1.1 Number of sampled adolescents	157	150	156	120	120	158	131	119	141
	1.2 Sum of weights	12,046	39,517	15,083	9,263	15,050	10,721	13,562	9,449	19,315
	1.3 CV	58.98	54.48	76.30	58.92	77.06	66.03	61.46	52.14	62.98
2	Nonresponse adjustment									
	2.1 Number of completed insurance interviews	137	130	142	107	109	144	107	109	124
	2.2 Sum of weights before adjustment	12,046	39,217	15,083	8,263	15,050	9,921	13,162	8,849	19,315
	a. Eligible respondents	10,706	33,866	13,624	8,218	13,080	9,767	11,141	8,796	17,017
	b. Ineligibles	42	34	59	2	73	73	37	0	78
	c. Nonrespondents	1,297	5,316	1,400	43	1,897	81	1,984	53	2,220
	2.3 Number of weights after adjustment	12,046	39,517	15,083	9,263	15,050	10,721	13,562	9,449	19,315
	2.4 Mean adjustment factor	1.13	1.16	1.11	1.01	1.15	1.02	1.18	1.01	1.14
3	Trimming adjustment**									
	3.1 Number of trimmed records	0	0	1	0	0	0	0	1	0
	3.2 Sum of weights before trimming adjustment	11,882	39,795	14,665	9,501	14,963	10,294	12,898	9,576	19,011
	3.3 Sum of weights after trimming adjustment	11,882	39,795	14,452	9,501	14,963	10,294	12,898	9,449	19,011
4	Raking nonresponse adjustment									
	4.1 Number of completed insurance interviews	136	133	141	108	109	144	106	110	122
	4.2 Sum of weights after adjustment	12,290	41,543	13,509	9,150	13,403	10,582	14,122	10,715	15,284
	4.3 Mean adjustment factor	1.03	1.06	0.90	1.00	0.90	1.03	1.08	1.13	0.80
	4.4 CV	81.90	69.79	84.08	77.10	82.30	80.39	85.99	55.24	74.50
	4.5 Mean weight	90.37	312.35	95.81	84.72	122.96	73.49	133.23	97.41	125.28

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

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Appendix B

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

	Categorical variables					Continuous variables				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
County/strata	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
State Total	1.62	1.59	2.56	0.38	1.26	1.98	1.86	2.94	1.51	1.40
Los Angeles	1.39	1.36	2.29	0.49	1.17	1.42	1.33	2.06	1.10	1.18
Long Beach	1.37	1.41	2.01	0.48	1.16	1.39	1.32	1.89	0.89	1.17
Pasadena	1.49	1.51	2.20	0.48	1.21	1.19	1.10	1.58	1.04	1.09
Remainder of Los Angeles	1.34	1.32	2.17	0.50	1.15	1.38	1.31	1.92	1.14	1.17
San Diego	1.24	1.23	1.89	0.46	1.11	1.32	1.39	1.55	0.92	1.14
Orange	1.31	1.31	1.93	0.48	1.14	1.42	1.36	2.02	1.02	1.18
Santa Clara	1.34	1.31	2.15	0.46	1.15	1.48	1.50	1.76	1.11	1.21
San Bernardino	1.21	1.21	2.05	0.40	1.09	1.35	1.32	1.90	1.02	1.15
Riverside	1.24	1.28	1.65	0.37	1.10	1.38	1.44	1.67	0.87	1.17
Alameda	1.86	1.88	2.88	0.68	1.35	2.62	2.63	3.63	1.79	1.60
Berkeley	2.35	2.33	3.98	0.94	1.51	2.46	2.33	3.44	1.50	1.55
Remainder of Alameda	1.27	1.35	1.97	0.45	1.12	1.72	1.77	2.27	1.23	1.30
Sacramento	1.23	1.29	1.59	0.59	1.10	1.26	1.23	1.63	0.97	1.12
Contra Costa	1.26	1.33	2.11	0.43	1.11	1.66	1.47	2.91	1.15	1.27
Fresno	1.47	1.49	2.55	0.44	1.20	1.58	1.59	2.13	1.12	1.25
San Francisco	1.42	1.51	1.82	0.52	1.18	1.47	1.44	1.79	1.22	1.21
Ventura	1.37	1.35	2.32	0.42	1.16	1.43	1.31	2.29	1.14	1.19
San Mateo	1.31	1.36	1.79	0.45	1.14	1.68	1.61	2.87	0.95	1.27
Kern	1.20	1.20	1.89	0.45	1.09	1.35	1.24	2.03	0.78	1.15
San Joaquin	1.34	1.28	2.40	0.41	1.14	1.36	1.34	1.74	1.06	1.16
Sonoma	1.23	1.20	2.09	0.44	1.10	1.42	1.51	1.89	0.73	1.18
Stanislaus	1.34	1.32	3.85	0.56	1.14	1.38	1.43	1.68	1.02	1.17
Santa Barbara	1.43	1.41	2.49	0.48	1.18	1.29	1.28	1.64	0.92	1.13
Solano	1.36	1.30	1.94	0.50	1.16	1.22	1.09	1.72	0.97	1.10
Tulare	1.48	1.49	2.66	0.61	1.20	1.35	1.33	1.65	1.11	1.16
Santa Cruz	1.24	1.23	1.95	0.62	1.11	1.27	1.22	1.60	1.04	1.12
Marin	1.40	1.41	2.80	0.66	1.17	1.28	1.22	1.61	0.85	1.13
San Luis Obispo	1.21	1.19	2.02	0.38	1.09	1.33	1.45	1.64	0.68	1.14
Placer	1.15	1.15	2.02	0.37	1.06	1.26	1.23	1.81	0.92	1.11
Merced	1.31	1.34	2.04	0.55	1.14	1.46	1.49	1.73	1.15	1.20
Butte	1.26	1.37	1.75	0.42	1.11	1.64	1.29	3.20	0.97	1.25

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

	Categorical variables					Continuous variables				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
County/strata	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
Shasta	1.34	1.29	2.28	0.51	1.14	1.38	1.24	2.43	0.97	1.16
Yolo	1.39	1.41	2.76	0.52	1.16	1.36	1.37	1.89	0.91	1.16
El Dorado	1.41	1.39	1.94	0.37	1.18	1.27	1.32	1.57	0.82	1.12
Imperial	1.28	1.26	2.00	0.38	1.12	1.40	1.42	1.71	0.96	1.18
Napa	1.32	1.32	1.91	0.44	1.14	1.38	1.15	2.29	0.79	1.15
Kings	1.31	1.33	2.23	0.51	1.13	1.89	1.85	2.57	1.23	1.36
Madera	1.35	1.34	2.05	0.46	1.15	1.67	1.63	2.10	1.36	1.29
Monterey, San Benito	1.27	1.36	2.18	0.46	1.11	1.30	1.55	1.65	0.68	1.12
Del Norte, Humboldt	1.24	1.29	1.93	0.41	1.11	1.49	1.45	2.13	1.17	1.22
Lassen, Modoc, Siskiyou, Trinity	1.27	1.29	2.08	0.52	1.12	1.58	1.48	2.86	0.79	1.23
Lake, Mendocino	1.22	1.28	1.55	0.42	1.10	1.46	1.42	1.93	0.93	1.20
Colusa, Glen, Tehama	1.28	1.31	1.94	0.31	1.12	1.68	1.33	3.63	1.09	1.26
Sutter, Yuba	1.23	1.21	1.62	0.55	1.10	1.29	1.27	1.74	0.96	1.13
Plumas, Nevada, Sierra	1.27	1.28	1.71	0.45	1.11	1.65	1.32	3.21	1.03	1.26
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.20	1.26	1.83	0.35	1.09	1.33	1.22	1.95	1.05	1.15

Note: Twenty-eight categorical variables and 7 continuous variables were used in the adult table.

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

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

	Categorical variables					Continuous variables				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
County/strata	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
State Total	1.78	1.90	2.31	0.44	1.32	1.92	1.96	2.11	1.63	1.38
Los Angeles	1.46	1.52	2.31	0.43	1.19	1.68	1.66	1.92	1.42	1.29
Long Beach	1.29	1.38	2.06	0.27	1.12	1.26	1.24	1.84	0.78	1.11
Pasadena	1.48	1.28	3.76	0.41	1.18	1.79	1.62	3.10	1.11	1.32
Remainder of Los Angeles	1.38	1.43	2.28	0.44	1.16	1.63	1.65	1.87	1.29	1.27
San Diego	1.47	1.36	2.82	0.37	1.19	1.47	1.40	1.95	1.16	1.21
Orange	1.41	1.29	3.42	0.56	1.16	1.45	1.45	1.72	1.21	1.20
Santa Clara	1.47	1.41	3.22	0.22	1.17	1.47	1.50	1.83	1.02	1.21
San Bernardino	1.60	1.67	2.83	0.32	1.24	1.72	1.56	2.37	1.44	1.30
Riverside	1.46	1.47	2.52	0.21	1.18	1.42	1.43	1.93	1.05	1.19
Alameda	1.95	2.02	3.25	0.50	1.37	1.89	1.90	2.22	1.56	1.37
Berkeley	2.08	2.28	3.34	0.08	1.38	2.55	2.27	4.26	1.90	1.58
Remainder of Alameda	1.53	1.60	2.49	0.40	1.22	1.51	1.51	1.82	1.24	1.23
Sacramento	1.21	1.19	2.33	0.18	1.08	1.43	1.50	1.78	0.80	1.19
Contra Costa	1.29	1.36	1.96	0.10	1.11	1.33	1.36	1.52	1.15	1.15
Fresno	1.68	1.58	3.30	0.21	1.26	1.70	1.67	2.12	1.18	1.30
San Francisco	1.43	1.44	2.35	0.67	1.17	1.47	1.40	2.25	1.06	1.21
Ventura	1.41	1.43	2.31	0.18	1.16	1.43	1.49	1.59	1.13	1.19
San Mateo	1.30	1.39	1.89	0.11	1.10	1.37	1.43	1.57	0.92	1.17
Kem	1.32	1.29	2.32	0.15	1.12	1.46	1.32	2.01	1.20	1.20
San Joaquin	1.33	1.56	1.84	0.15	1.13	1.34	1.31	1.76	1.00	1.15
Sonoma	1.19	1.35	1.72	0.01	1.05	1.39	1.41	2.11	0.58	1.16
Stanislaus	1.10	1.14	1.88	0.11	1.02	1.53	1.51	1.66	1.42	1.24
Santa Barbara	1.33	1.34	1.94	0.09	1.12	1.38	1.33	1.90	1.07	1.17
Solano	1.37	1.33	2.36	0.34	1.15	1.31	1.19	2.15	1.00	1.13
Tulare	1.32	1.37	1.78	0.34	1.13	1.47	1.38	2.01	0.99	1.20
Santa Cruz	1.32	1.30	1.87	0.08	1.12	1.40	1.32	1.80	1.05	1.18
Marin	0.94	0.95	1.66	0.16	0.94	1.51	1.34	2.33	1.16	1.22
San Luis Obispo	0.97	1.11	1.41	0.14	0.96	1.18	1.18	1.50	0.97	1.08
Placer	1.21	1.21	1.80	0.42	1.09	1.17	1.15	1.53	0.98	1.08
Merced	1.45	1.57	1.85	0.06	1.17	1.75	1.60	3.00	0.94	1.30
Butte	1.25	1.33	1.83	0.31	1.09	1.41	1.48	1.83	0.93	1.18
Shasta	1.36	1.44	1.89	0.01	1.13	1.28	1.24	1.65	1.13	1.13
Yolo	1.23	1.11	2.35	0.01	1.06	1.47	1.41	1.93	1.08	1.21
El Dorado	1.34	1.32	2.32	0.08	1.11	1.37	1.43	1.70	0.99	1.16
Imperial	1.44	1.34	4.03	0.41	1.17	1.18	1.24	1.43	0.86	1.08

Table B-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
Napa	1.52	1.29	3.23	0.56	1.20	1.34	1.32	1.67	1.00	1.15
Kings	1.38	1.40	2.16	0.25	1.16	1.54	1.51	2.18	1.11	1.23
Madera	1.16	1.18	1.67	0.22	1.05	1.44	1.48	1.74	1.14	1.20
Monterey, San Benito	1.57	1.47	3.47	0.25	1.22	1.60	1.61	1.97	1.21	1.26
Del Norte, Humboldt	1.65	1.60	2.88	0.10	1.25	1.50	1.23	2.66	1.05	1.21
Lassen, Modoc, Siskiyou, Trinity	1.30	1.42	2.26	0.15	1.11	1.44	1.19	2.41	0.94	1.18
Lake, Mendocino	1.34	1.42	1.93	0.02	1.12	1.31	1.31	1.52	1.09	1.14
Colusa, Glen, Tehama	1.17	1.26	1.62	0.22	1.06	1.29	1.24	1.66	1.03	1.13
Sutter, Yuba	1.35	1.37	2.48	0.07	1.12	1.45	1.43	1.80	1.18	1.20
Plumas, Nevada, Sierra	1.10	1.14	1.61	0.04	1.01	1.25	1.29	1.42	1.00	1.11
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.31	1.36	1.80	0.52	1.13	1.43	1.44	1.75	1.02	1.19

Note: Seventeen categorical variables and 6 continuous variables were used in the adult table.

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

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

	Categorical variables					Continuous variables				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
County/strata	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
State Total	1.89	1.96	2.51	1.12	1.37	1.94	2.06	2.39	1.03	1.38
Los Angeles	1.53	1.51	2.17	0.83	1.23	1.76	1.75	2.27	1.22	1.32
Long Beach	1.61	1.66	1.82	1.18	1.26	1.66	1.55	2.47	1.07	1.28
Pasadena	1.16	1.18	1.69	0.32	1.05	0.91	0.88	1.16	0.64	0.95
Remainder of Los Angeles	1.54	1.49	2.21	0.85	1.23	1.73	1.76	2.35	1.14	1.31
San Diego	1.41	1.32	1.74	1.07	1.18	1.25	1.23	1.81	0.71	1.11
Orange	1.40	1.45	1.82	1.06	1.18	1.16	1.25	1.70	0.30	1.05
Santa Clara	1.37	1.37	2.13	0.80	1.16	1.20	1.20	1.90	0.77	1.08
San Bernardino	1.59	1.53	2.38	1.26	1.26	1.45	1.20	2.62	0.94	1.18
Riverside	1.42	1.34	2.34	1.05	1.19	1.41	1.07	2.75	0.82	1.16
Alameda	2.05	2.05	2.66	0.72	1.42	1.73	1.78	2.36	0.97	1.31
Berkeley	1.52	1.62	3.05	0.39	1.18	1.88	1.16	3.91	0.92	1.31
Remainder of Alameda	1.68	1.82	2.18	0.58	1.28	1.43	1.53	1.86	0.75	1.18
Sacramento	1.55	1.43	2.37	0.88	1.23	1.42	1.30	1.83	1.17	1.19
Contra Costa	1.13	1.12	1.83	0.76	1.06	1.22	1.42	1.65	0.51	1.09
Fresno	1.59	1.54	2.67	0.91	1.25	1.47	1.59	2.25	0.86	1.20
San Francisco	1.73	1.67	3.04	0.32	1.28	1.18	1.27	1.53	0.53	1.07
Ventura	1.40	1.54	1.92	0.78	1.17	1.28	1.29	2.04	0.47	1.11
San Mateo	1.27	1.35	1.90	0.27	1.10	1.36	1.33	1.87	1.02	1.16
Kern	1.51	1.54	1.96	1.08	1.23	1.40	1.28	2.24	0.90	1.17
San Joaquin	1.42	1.49	2.13	0.77	1.18	1.42	1.37	1.77	1.14	1.19
Sonoma	1.71	1.63	2.84	0.96	1.29	1.16	1.07	2.35	0.35	1.04
Stanislaus	1.31	1.23	2.17	0.82	1.13	1.21	1.12	1.72	0.75	1.09
Santa Barbara	1.41	1.32	2.10	0.91	1.18	1.37	1.50	2.06	0.44	1.15
Solano	1.44	1.44	2.07	0.79	1.19	1.54	1.51	2.34	1.16	1.23
Tulare	1.66	1.59	2.36	0.66	1.27	1.97	1.88	3.95	0.84	1.36
Santa Cruz	1.54	1.55	2.07	1.21	1.24	1.34	1.17	1.91	0.99	1.15
Marin	1.22	1.25	1.43	0.92	1.10	1.09	1.11	1.35	0.77	1.04
San Luis Obispo	1.22	1.26	1.77	0.63	1.09	1.35	1.31	1.71	1.12	1.16
Placer	2.10	2.21	3.31	0.94	1.42	1.51	1.40	2.76	0.79	1.21
Merced	1.82	1.89	2.46	1.04	1.34	1.53	1.47	2.55	0.50	1.20
Butte	1.16	1.08	1.82	0.66	1.07	1.13	1.01	2.18	0.31	1.03
Shasta	1.61	1.56	2.53	0.92	1.26	1.26	1.20	1.72	0.78	1.11
Yolo	1.16	1.17	1.51	0.63	1.07	1.50	1.45	1.92	1.13	1.22
El Dorado	1.68	1.65	2.17	0.87	1.29	1.16	1.07	1.85	0.68	1.07
Imperial	1.29	1.31	1.61	1.01	1.13	1.37	1.52	1.79	0.63	1.16

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

	Categorical variables					Continuous variables				
	Design effect (DEFF)				DEFT	Design effect (DEFF)				DEFT
County/strata	Average	Median	Maximum	Minimum	Average	Average	Median	Maximum	Minimum	Average
Napa	1.61	1.60	2.17	0.91	1.26	1.15	1.24	1.39	0.77	1.06
Kings	1.73	1.51	3.63	0.69	1.29	1.44	1.50	2.13	0.61	1.18
Madera	1.45	1.30	2.61	0.82	1.19	1.29	1.29	1.66	0.94	1.13
Monterey, San Benito	1.47	1.45	2.09	0.74	1.20	1.31	1.29	1.84	0.86	1.14
Del Norte, Humboldt	1.34	1.34	2.63	0.51	1.14	1.35	1.34	1.66	1.07	1.16
Lassen, Modoc, Siskiyou, Trinity	1.44	1.52	2.21	0.91	1.19	1.53	1.50	1.79	1.17	1.23
Lake, Mendocino	1.75	1.82	2.48	0.39	1.30	1.43	1.37	2.05	0.76	1.18
Colusa, Glen, Tehama	1.59	1.50	2.76	0.69	1.24	1.08	1.16	1.29	0.75	1.03
Sutter, Yuba	1.38	1.46	1.97	0.50	1.16	1.38	1.39	2.22	0.66	1.16
Plumas, Nevada, Sierra	1.28	1.20	1.67	0.95	1.12	0.99	0.90	1.29	0.77	0.99
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	1.60	1.50	2.65	1.03	1.25	1.85	1.59	3.49	0.93	1.33

Note: Seventeen categorical variables and 6 continuous variables were used in the adult table.

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Appendix C

Table C-1. Race-ethnic supplemental samples

	Korean	Japanese	Cambodian	Vietnamese	Shasta Latino	AIAN
1. Household weights						
1.1 Base weight						
a. Number of sampled records	348,771	347,599	347,701	348,119	347,039	348,089
b. Sum of Design base weight	27,445,848	27,445,743	27,445,659	27,445,825	27,445,650	27,976,878
c Coefficient of variation	51.21	51.48	52.17	51.06	52.02	52.26
d. Sum of Model base weight	348,771	347,599	347,701	348,119	347,039	348,089
e. Coefficient of variation	0.00	0.00	0.00	0.00	0.00	0.00
1.2 Subsampling Adjustment						
a. Number of sampled records	308,463	307,291	307,393	307,811	306,731	307,781
b. Sum of Design weights after adjustment	27,445,848	27,445,743	27,445,659	27,445,825	27,445,650	27,976,878
c. Coefficient of variation	53.34	53.60	54.30	53.13	54.12	54.21
d. Sum of Model weights after adjustment	348,771	347,599	347,701	348,119	347,039	348,089
e. Coefficient of variation	11.13	11.12	11.12	11.12	11.12	11.12
1.3 Multiple Telephone Adjustment						
a. Number of sampled records	811	903	207	880	339	3,186
b. Sum of Design weights after adjustment	66,221	63,011	7,077	74,350	1,755	237,311
c. Coefficient of variation	44.92	58.29	127.70	48.07	238.23	65.72
1.4 Race-ethnic adjustment						
a. Number of sampled records	811	903	207	880	339	3,186
b. Sum of Design weights after adjustment	67,004	66,090	7,153	78,135	1,755	344,520
c. Coefficient of variation	46.11	61.93	128.17	56.00	238.23	75.33
d. Race-ethnic adjustment Factor	1.41	1.34	1.08	2.05	0.00	1.26
e. Sum of Model weights after adjustment	865	1,010	229	966	341	4,616
f. Coefficient of variation	15.07	22.77	34.14	33.25	4.01	23.98
g. Race-ethnic adjustment Factor	1.65	1.67	2.67	2.67	0.00	1.37

Table C-1. Race-ethnic supplemental samples (continued)

	Korean	Japanese	Cambodian	Vietnamese	Shasta Latino	AIAN
2. Person weights						
2.1 Creation of Adult weights						
a. Number of sampled records	811	903	207	880	339	3,186
b. Sum of Design Adult weights	136,193	124,795	18,810	194,177	3,052	667,108
c. Coefficient of variation	71.54	95.12	160.44	76.38	250.85	99.22
d. Sum of Model Adult weights	1,754	1,846	584	2,420	553	9,025
e. Coefficient of variation	54.58	64.15	64.96	59.54	57.06	59.31
2.2 Creation of Child weights						
a. Number of sampled records	205	155	61	227	115	755
b. Sum of Design Child weights	28,996	20,968	3,621	42,783	928.92	172,121
c. Coefficient of variation	70.41	97.55	180.89	96.27	308.74	115.16
d. Sum of Model Child weights	370	320	154	532	224	2,254
e. Coefficient of variation	52.69	71.08	63.57	79.40	48.76	73.11
2.3 Creation of adolescent weights						
a. Number of sampled records	77	70	54	73	51	368
b. Sum of Design Adolescent weights	8,726	7,724	2,541	12,242	189	61,678
c. Coefficient of variation	60.25	86.85	145.69	72.15	241.80	104.49
d. Sum of Model Adolescent weights	105	112	123	141	77	874
e. Coefficient of variation	41.56	54.08	47.78	54.01	55.39	64.97
3. Final weights						
3.1 Adult final weight						
a. Number of sampled records	811	903	207	880	339	3186
b. Sum of Design Adult final weights	278,912	298,697	48,680	346,824	5,136	413,493
c. Coefficient of variation	77.00	93.42	163.20	88.15	257.96	111.26
d. Sum of Model Adult final weights	278,912	298,698	48,680	346,824	5,136	413,494
e. Coefficient of variation	61.54	61.90	76.87	70.01	72.0883	84.77

Table C-1. Race-ethnic supplemental samples (continued)

	Korean	Japanese	Cambodian	Vietnamese	Shasta Latino	AIAN
3.2 Child final weight						
a. Number of sampled records	204	130	59	218	106	416
b. Sum of Design Child final weights	57,671	62,974	19,784	86,727	2,435	133,090
c. Coefficient of variation	76.33	81.03	180.62	77.64	310.70	119.60
d. Sum of Model Child final weights	57,671	62,974	19,784	86,727	2,433	133,090
e. Coefficient of variation	58.52	62.39	74.22	64.91	50.4881	87.99
3.3 Adolescent final weight						
a. Number of sampled records	76	61	52	71	45	165
b. Sum of Design Adolescent final weights	32,681	26,757	15,168	45,321	1,187	67,641
c. Coefficient of variation	65.84	73.15	184.43	78.23	221.51	80.48
d. Sum of Model Adolescent final weights	32,681	26,757	15,168	45,321	1,189	67,640
e. Coefficient of variation	49.60	54.55	55.81	64.08	60.6455	100.46
3.4 Design-based overall raking factor						
a. Adults	2.05	2.39	2.59	1.79	1.68	0.62
b. Children	1.99	3.00	5.46	2.03	2.62	0.77
c. Adolescents	3.75	3.46	5.97	3.70	6.27	1.10

Table C-1. Race-ethnic supplemental samples (continued)

	Bangladeshi	Indian	Pakistani	Sri Lankan	South Asian
1. Household weights					
1.1 Base weight					
a. Number of sampled records					348,806
b. Sum of Design base weight					27,445,731
c Coefficient of variation					52.12
d. Sum of Model base weight					348,806
e. Coefficient of variation					0.00
1.2 Subsampling Adjustment					
a. Number of sampled records					308,498
b. Sum of Design weights after adjustment					27,445,731
c. Coefficient of variation					54.30
d. Sum of Model weights after adjustment					348,806
e. Coefficient of variation					11.13
1.3 Multiple Telephone Adjustment					
a. Number of sampled records					846
b. Sum of Design weights after adjustment					46,913
c. Coefficient of variation					95.78
1.4 List Adjustment					
a. Number of sampled records					846
b. Sum of Design weights after adjustment					47,342
c. Coefficient of variation					96.07
d. List Adjustment Factor					1.18
e. Sum of Model weights after adjustment					903
f. Coefficient of variation					20.30
g. List Adjustment Factor					2.07

Table C-1. Race-ethnic supplemental samples (continued)

	Bangladeshi	Indian	Pakistani	Sri Lankan	South Asian
2. Person weights					
2.1 Creation of Adult weights					
a. Number of sampled records					846
b. Sum of Design Adult weights					102,761
c. Coefficient of variation					114.21
d. Sum of Model Adult weights					1,950
e. Coefficient of variation					54.40
2.2 Creation of Child weights					
a. Number of sampled records					272
b. Sum of Design Child weights					25,184
c. Coefficient of variation					125.97
d. Sum of Model Child weights					468
e. Coefficient of variation					62.94
2.3 Creation of adolescent weights					
a. Number of sampled records					77
b. Sum of Design Adolescent weights					6,967
c. Coefficient of variation					102.07
d. Sum of Model Adolescent weights					143
e. Coefficient of variation					66.16
3. Final weights					
3.1 Adult final weight					
a. Number of sampled records	14	761	65	11	846
b. Sum of Design Adult final weights	2,852.56	261,396.00	18,222.00	5,321.82	286,259
c. Coefficient of variation	102.83	131.66	105.69	50.42	129.59
d. Sum of Model Adult final weights	3,248.86	261,395.99	18,222.00	5560	286,516
e. Coefficient of variation	56.87	76.63	55.88	25.88	75.32

Table C-1. Race-ethnic supplemental samples (continued)

	Bangladeshi	Indian	Pakistani	Sri Lankan	South Asian
3.2 Child final weight					
a. Number of sampled records	5	232	23	3	257
b. Sum of Design Child final weights	1,200.44	67,672.42	6,820.45	1,107.06	74,326
c. Coefficient of variation	116.84	124.97	153.02	73.58	126.56
d. Sum of Model Child final weights	804.14	67,672.33	6,072.03	1,074.48	72,466
e. Coefficient of variation	51.70	75.21	95.85	16.72	63.52
3.3 Adolescent final weight					
a. Number of sampled records	0	60	9	1	69
b. Sum of Design Adolescent final weights	0.00	26,745.58	2,483.55	640.12	28,863
c. Coefficient of variation	0.00	98.29	109.27	NA	99.24
d. Sum of Model Adolescent final weights	0.00	26,745.68	3,231.97	434.15	29,833
e. Coefficient of variation	0.00	42.86	92.92	NA	48.90
Design-based overall raking factor					
a. Adults					2.79
b. Children					2.95
c. Adolescents					4.14

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Table C-2. Sample size, sums of weights and coefficient of variation for the race-ethnic samples

								AIAN			
			Japanese	Cambodian	Korean	Vietnamese	South Asian	Urban	Rural	All	Shasta
Design-based	All	n	798	197	791	843	822	799	773	1,572	339
		SUM	257,669	42,701	261,761	328,085	261,216	140,072	101,596	241,667	5,136
		CV	94.79	172.11	71.97	88.83	121.93	101.91	130.55	114.83	257.96
	List	n	330	126	326	540	439	248	103	351	301
		SUM	54,633	3,681	72,589	161,622	40,368	28,731	8,302	37,032	928
		CV	48.29	66.38	54.11	85.30	59.02	76.04	159.40	97.33	86.79
	RDD	n	468	71	465	303	383	551	670	1,221	38
		SUM	203,036	39,020	189,172	166,463	220,848	111,341	93,294	204,635	4,208
		CV	81.68	84.04	66.15	76.51	76.29	99.60	126.44	113.59	52.07
Model-based	All	n	798	197	791	843	822	799	773	1,572	339
		SUM	255,677	41,755	260,770	326,491	262,321	137,274	103,295	240,569	5,136
		CV	58.16	60.31	51.55	69.69	55.97	89.33	82.53	88.41	72.09
	List	n	330	126	326	540	439	248	103	351	301
		SUM	90,331	27,032	100,781	199,984	136,255	21,658	7,043	28,701	4,411
		CV	50.49	59.03	51.69	71.93	53.57	66.16	75.66	69.30	74.53
	RDD	n	468	71	465	303	383	551	670	1,221	38
		SUM	165,346	14,722	159,989	126,506	126,066	115,616	96,253	211,869	725
		CV	58.81	63.02	51.00	65.60	58.18	79.77	78.97	83.06	53.63

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Table C-3. Variables used in the evaluation of the weights

Variable	Description
Gender	1. Male 2. Female
Age	1. 18 to 25 2. 26 to 30 3. 31 to 37 4. 38 to 45 5. 46 to 53 6. 54 to 64 7. 65 plus
General Health	1. Excellent 2. Very Good 3. Good 4. Fair or Poor
Energy	1. All the time 2. Most of the time 3. Some of the time 4. A little or not at all
Feel downhearted and sad	1. At least some of the time 2. A little of the time 3. Never
Firearms in home	1. Yes 2. No
Alcoholic beverages. in last month?	1. Yes 2. No
Smoked >=100 cigarettes?	1. Yes 2. No
Walked or Biked to work?	1. Yes 2. No
Strength building exercises in the past month?	1. Yes 2. No
Diagnosed with cancer?	1. Yes 2. No
How often wear a long sleeved shirt when in the sun sun?	1. Always 2. Sometimes 3. Never
Family member ever had cancer?	1. Yes 2. No
Last visit to dentist etc.	1. Never 2. 1 - 6 mo. Ago 3. 7 - 12 mo. Ago 4. 1 - 2 yrs. Ago 5. 2 - 5 yrs. Ago 6. more than 5 yrs.
Have dental insurance?	1. Yes 2. No
Place to go when sick?	1. Yes 2. No

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Table C-3. Variables used in the evaluation of the weights (continued)

Variable	Description
Type of place you go to when sick.	1. Dr./Kaiser/HMO 2. Clinic/Center/Hosp. 3. ER/Other/No place
Felt discriminated against regarding health care?	1. Yes 2. No
In what country were you born?	1. U.S. 3. Cambodia 6. Latino Country 13. S. Asian country 15. Japan 16. Korea 24. Vietnam 91. Other
In what country was your mother born?	1. U.S. 3. Cambodia 6. Latino Country 13. S. Asian country 15. Japan 16. Korea 24. Vietnam 91. Other
In what country was your father born?	1. U.S. 3. Cambodia 6. Latino Country 13. S. Asian country 15. Japan 16. Korea 24. Vietnam 91. Other
Languages spoken at home	1. English only 2. English and other 3. No English
Are you a citizen?	1. Yes 2. No 3. App. Pending
Type of clinic (see AH3)	1. HMO/Kaiser/Dr. office 2. County or govt clinic 3. Hospital/Outpatient 4. VA/Veterans hosp 5. ER 6. Urgent care clinic 7. Chiropractic clinic 8. Indian Health Services 9. School clinic 91. Other
Permanent resident with a green card?	1. Yes 2. No

Table C-3. Variables used in the evaluation of the weights (continued)

Variable	Description
Number of years in the U.S.	1. 1 to 5 years 2. 6 to 20 3. 21 plus
Marriage status	1. Married 2. Living with partner 3. Widowed 4. Divorced 5. Separated 6. Never married
Education	1. HS or less 2. Some college 3. BS Degree or advanced
Currently working for wages?	1. Yes 2. No 3. Calworks
Times seen a medical dr. in past 12 months.	0. None 1. Once 2. Twice 3. Three times 4. Four or more
Talk to some other health person?	1. Yes 2. No
Covered by some other government health plan?	1. Yes 2. No
Covered by HIS, THP or UIC?	1. Yes 2. No
Annual HH income over $20K?	1. Yes 2. No
Receiving AFDC, TANF or CalWORKS?	1. Yes 2. No
Receiving public housing subsidies?	1. Yes 2. No
Cut meal size in past 12 mo. because no money?	1. Yes 2. No

Source: UCLA Center for Health Policy Research, 2001 California Health Interview Survey.

Note: Table C-2 to C-10 are in the file *ADULT OVERSAMPLE SUMMARY (DELIVERY).XLS* attached to this memo. These tables are best printed on legal size paper.

The California Health Interview Survey (CHIS) is a collaboration of:

UCLA Center for
Health Policy Research

California Department
of Health Services

Public Health
Institute

California Health Interview Survey (CHIS)
UCLA Center for Health Policy Research
10911 Weyburn Avenue, Suite 300
Los Angeles, California 90024-2887
Phone 310-794-0925
Toll Free 1-866-275-2447
Fax 310-794-2686
chis@ucla.edu
www.chis.ucla.edu