



*california
health
interview
survey*

Report One

CHIS 2005 Methodology Series

Sample Design

CALIFORNIA HEALTH INTERVIEW SURVEY

CHIS 2005 METHODOLOGY SERIES

REPORT 1

SAMPLE DESIGN

April 2007

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



www.chis.ucla.edu

This report provides analysts with information about the sampling methods used for CHIS 2005, including both the household and person (within household) sampling. This report also provides a discussion on achieved sample size and how it compares to the planned sample size.

Suggested citation:

California Health Interview Survey. *CHIS 2005 Methodology Series: Report 1 - Sample Design*. Los Angeles, CA: UCLA Center for Health Policy Research, 2007.

Copyright © 2007 by the Regents of the University of California.

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 2005 came from multiple sources: the California Department of Health Services, the California Endowment, the National Cancer Institute, the First 5 California, the Robert Wood Johnson Foundation, the California Department of Mental Health, the California Office of the Patient Advocate, Kaiser Permanente, the San Diego County Human and Health Services Agency, the Marin County Department of Human and Health Services, the First 5 Marin Children and Families Commission, the Center for Public Policy Research at the University of California, Davis, the U. S. Centers for Disease Control and Prevention, the Solano County Health and Social Services Department, and the Humboldt County Department of Health and Human Services.

PREFACE

Sample Design is the first in a series of methodological reports describing the 2005 California Health Interview Survey (CHIS 2005). 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 for the 2005 survey. The survey examines public health and health care access issues in California. The CHIS telephone survey is the largest state health survey ever undertaken in the United States. The plan is to monitor the health of Californians and examine changes over time by conducting periodic surveys in the future.

Methodological Reports

The first five methodological reports for CHIS 2005 are as follows:

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

This report describes the procedures used to design and select the sample from CHIS 2005. An appropriate sample design is a feature of a successful survey, and CHIS 2005 presented many issues that had to be addressed at the design stage. This report explains why the design features of CHIS were selected and presents the alternatives that were considered.

The primary purpose of this report is to provide analysts information about the sampling methods used for CHIS 2005, including both the household and person (within household) sampling. In general terms, once a household was sampled, an adult within that household was sampled. If there were

children and/or adolescents in the household, one child and/or one adolescent was eligible for sampling. This report also provides a discussion on achieved sample size and how it compares to the planned sample size.

TABLE OF CONTENTS

<u>Chapter</u>		<u>Page</u>
	PREFACE	i
1	CHIS 2005 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 Rates	1-4
	1.5 Weighting the Sample.....	1-8
	1.6 Imputation Methods.....	1-10
	1.7 Methodology Report Series	1-11
2	TELEPHONE SAMPLING METHODS.....	2-1
	2.1 List-Assisted Random-Digit-Dial Sampling.....	2-1
	2.2 Noncoverage Issues in Telephone Surveys.....	2-3
	2.3 Supplemental Sampling	2-4
3	SAMPLING HOUSEHOLDS	3-1
	3.1 Population of Interest.....	3-1
	3.2 Sample Design	3-1
	3.2.1 Base RDD Sample.....	3-2
	3.2.2 Supplemental Geographic Samples.....	3-5
	3.2.3 Supplemental Surname List Samples	3-5
	3.2.4 Supplemental Child Samples	3-6
	3.3 Stratification	3-8
	3.4 Sample Selection.....	3-11
	3.5 Expected Design Effect	3-12
4	WITHIN-HOUSEHOLD SAMPLING.....	4-1
	4.1 Sampling Alternatives.....	4-1
	4.2 Child First Procedure	4-2
	4.3 Adult Sampling	4-5
	4.4 Child Sampling	4-6
	4.5 Adolescent Sampling	4-7
5	ACHIEVED SAMPLE SIZES.....	5-1
	REFERENCES	R-1

TABLE OF CONTENTS (continued)

APPENDIX.....	A-1
---------------	-----

List of Tables

<u>Table</u>		<u>Page</u>
1-1	California county and county group strata used in the CHIS 2005 sample design.....	1-2
1-2	Number of completed CHIS 2005 interviews by type of sample and instrument	1-4
1-3	CHIS 2005 survey topic areas by instrument	1-5
2-1	CSSR result codes and their distribution in the CHIS 2005 sample	2-2
3-1	Targeted number of complete adult and child interviews for the base RDD sample by county	3-3
3-2	Targeted number of interviews for the supplemental geographic samples	3-5
3-3	Targeted number of complete adult interviews for the Korean and Vietnamese supplemental list samples.....	3-6
3-4	Number of records in the surname frames	3-6
3-5	Targeted number of complete adult and child interviews for child supplemental samples	3-7
3-6	Targeted number of complete adult and child interviews for San Diego County.....	3-7
3-7	Targeted number of complete adult and child interviews for all samples	3-8
3-8	Definition of sampling substrata, number of exchanges, and total number of households for Los Angeles County, San Diego County, Orange County, and Santa Clara County	3-10
3-9	Number of telephone numbers drawn by type of sample	3-12
3-10	Expected design effects and effective adult sample size associated with the sample allocation for the base RDD sample	3-14

TABLE OF CONTENTS (continued)

List of Tables (continued)

<u>Table</u>		<u>Page</u>
4-1	Effect of the child-first procedure on completed child and adolescent interviews.....	4-4
4-2	Distribution of households with children by type of child sampling	4-7
5-1	Number of completed adult interviews for the base RDD and geographic samples by sampling and self-reported stratum	5-2
5-2	Number of completed child interviews for the base RDD and geographic samples by sampling and self-reported stratum	5-4
5-3	Number of completed adolescent interviews for the base RDD and geographic samples by sampling and self-reported stratum	5-6
5-4	Number of completed adult, child, and adolescent interviews by surname list sample	5-7
5-5	Number of completed adult and child interviews by type of child supplemental sample.....	5-8
A-1	Stratum definitions for CHIS 2001, 2003, and 2005	A-1
A-2	Number of telephone numbers drawn by sample type and sampling stratum ..	A-3
A-3	Number of adult completed interviews by self-reported stratum	A-5
A-4	Number of child completed interviews by self-reported stratum	A-7
A-5	Number of adolescents completed interviews by self-reported stratum	A-9

1. CHIS 2005 DESIGN AND METHODOLOGY SUMMARY

1.1 Overview

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

The sample is designed to meet and optimize two goals:

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

The CHIS sample is representative of California's noninstitutionalized population living in households.

This series of reports describes the methods used in collecting data for CHIS 2005, the third CHIS data collection cycle, which was conducted between July 2005 and April 2006. The previous CHIS cycles (2001 and 2003) are described in similar series, available at http://www.chis.ucla.edu/methods_main.html.

CHIS data and results are used extensively by federal and State agencies, local public health agencies and organizations, advocacy and community organizations, other local agencies, hospitals, community clinics, health plans, foundations, and researchers. The data are widely used for analyses and publications to assess public health and health care needs, to develop and advocate policies to meet those needs, and to plan and budget health care coverage and services.

1.2 Sample Design Objectives

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

Table 1-1 shows the 44 sampling strata for CHIS 2005. A sufficient number of adult interviews were allocated to each stratum to support the first sample design objective. The geographic stratification of the state was revised from the design used in previous CHIS cycles, increasing the number of individual counties from 33 to 41.

Table 1-1. California county and county group strata used in the CHIS 2005 sample design

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

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

The samples in Marin, Humboldt, and Solano Counties were enhanced with additional funding. Additional samples were also selected statewide and in San Diego County to increase the number of child interviews; telephone numbers selected in these two additional samples were screened to identify households with children ages 0 to 11. All supplemental samples were implemented with and incorporated into the original statewide RDD sample.

The main RDD CHIS sample size is sufficient to accomplish the second objective. To increase the precision of estimates for Koreans and Vietnamese, areas with relatively high concentrations of these groups were sampled at higher rates. These geographically targeted oversamples were supplemented by telephone numbers associated with group-specific surnames drawn from listed telephone directories to further increase the sample size for Koreans and Vietnamese.

1.3 Data Collection

To capture the rich diversity of the California population, interviews were conducted in five languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, and Korean. These languages were chosen based on analysis of 2000 Census data to identify the languages that would cover the largest number of Californians in the CHIS sample that either did not speak English or did not speak English well enough to otherwise participate.

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

Table 1-2. Number of completed CHIS 2005 interviews by type of sample and instrument

Type of sample	Adult	Child	Adolescent
Total RDD + supplemental cases	43,020	11,358	4,029
RDD			
Base plus county supplements	41,074	9,605	3,739
Statewide child supplement	525	511	84
San Diego child supplement	1,143	1,160	186
Supplemental samples:			
Korean	199	60	14
Vietnamese	79	22	6

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Interviews in all languages were administered using Westat’s computer-assisted telephone interviewing (CATI) system. The average adult interview took 35 minutes to complete. The average child and adolescent interviews took 15 minutes and 20 minutes, respectively. For “child-first” interviews, additional household information asked as part of the child interview averaged almost 8 minutes. Interviews in non-English languages generally took longer to complete. Just over 10 percent of the adult interviews were completed in a language other than English, as were 18 percent of all child (parent proxy) interviews and 7 percent of all adolescent interviews.

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

1.4 Response Rates

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

Table 1-3. CHIS 2005 survey topic areas by instrument

Health status	Adult	Teen	Child
General health status, height and weight	✓	✓	✓
Days missed from school due to health problems	✓	✓	✓
Health conditions	Adult	Teen	Child
Asthma	✓	✓	✓
Diabetes	✓	✓	
Heart disease, high blood pressure, epilepsy	✓		
Physical disability/need for special equipment	✓		
Parental concerns with child development, attention deficit disorder (ADD)			✓
Mental health	Adult	Teen	Child
Mental health status	✓		✓
Perceived need, use of mental health services	✓		✓
Emotional functioning		✓	
Health behaviors	Adult	Teen	Child
Dietary intake	✓	✓	✓
Physical activity and exercise	✓	✓	✓
Walking for transportation and leisure	✓		
Sedentary time		✓	✓
Body image		✓	
Flu Shot	✓		
Alcohol and tobacco use	✓	✓	
Drug use		✓	
Sexual behavior, STD testing, birth control practices	✓	✓	
Women's health	Adult	Teen	Child
Pap test screening, mammography screening, hormone replacement therapy	✓		
Emergency contraception		✓	
Pregnancy status	✓	✓	
Cancer history and prevention	Adult	Teen	Child
Cancer history of respondent and family history	✓		
Colon cancer screening, prostate cancer (PSA) test	✓		
Dental health	Adult	Teen	Child
Last dental visit			✓
Dental insurance coverage		✓	✓
Injury	Adult	Teen	Child
Serious injuries (frequency, cause)		✓	✓
Injury prevention behaviors (bike helmets, seatbelts)		✓	

Table 1-3. CHIS 2005 survey topic areas by instrument (Continued)

Food insecurity/hunger	Adult	Teen	Child
Availability of food in household over past 12 months	✓		
Food environment	Adult	Teen	Child
Quality of food stores in area, where does teen/child eat lunch and breakfast		✓	✓
School has vending machines		✓	
Access to and use of health care	Adult	Teen	Child
Usual source of care, visits to medical doctor	✓	✓	✓
Emergency room visits	✓	✓	✓
Delays in getting care (prescriptions, tests, treatment)	✓	✓	✓
Racial/ethnic discrimination in health care, MD discussed diet and exercise	✓		
Communication problems with doctor	✓	✓	✓
Ability and parental knowledge of teen contacting a doctor		✓	
Health insurance	Adult	Teen	Child
Current insurance coverage, spouse's coverage, who pays for coverage	✓	✓	✓
Health plan enrollment, characteristics of plan	✓	✓	✓
Whether employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months	✓	✓	✓
Reasons for lack of insurance	✓	✓	✓
Public program eligibility	Adult	Teen	Child
Household poverty level	✓		
Program participation (TANF, CalWorks, Public Housing, Food Stamps, SSI, SSDI, WIC)	✓	✓	✓
Assets, alimony/child support/social security/pension	✓		
Eligible for Medi-Cal and healthy families	✓	✓	✓
Reason for Medi-Cal nonparticipation among potential eligibles	✓	✓	✓
Neighborhood and housing	Adult	Teen	Child
Neighborhood safety	✓	✓	
Home ownership, number of rooms, amount of mortgage/rent	✓		
Parental involvement/adult supervision	Adult	Teen	Child
Parental presence after school, parental knowledge of teen's activities		✓	
Child's activities with family			✓

Table 1-3. CHIS 2005 survey topic areas by instrument (Continued)

Child care and school attendance	Adult	Teen	Child
Current child care arrangements	✓		✓
Paid child care			✓
First 5 California: receipt of parent kit and attitudes towards preschool			✓
Preschool/school attendance, public/private school		✓	✓
Employment	Adult	Teen	Child
Employment status, spouse's employment status	✓		
Work in last week, industry and occupation	✓		
Hours worked at all jobs	✓	✓	
Income	Adult	Teen	Child
Respondent's and spouse's earnings last month before taxes	✓		
Household income (annual before taxes)	✓		
Number of persons supported by household income	✓		
Respondent characteristics	Adult	Teen	Child
Age, gender, height, weight, education	✓	✓	✓
Race and ethnicity	✓	✓	✓
Marital status	✓		
Sexual orientation	✓	✓	
Citizenship, immigration status, country of birth, length of time in U.S., languages spoken at home, English language proficiency	✓	✓	✓

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

The CHIS 2005 screener completion rate was 49.8 percent and was higher for households that were sent the advance letter. The extended interview completion rate varied across the adult, child and adolescent interviews. Multiplying the screener and extended rates gives an overall response rate for each type of interview. At the household level, the percentage of households completing one or more of the extended interviews (adult, child, and/or adolescent) is a useful summary of the overall success of the study. For CHIS 2005, the household response rate was 29.5 percent (the product of the screener response rate and the completion rate at the household level of 59.3 %). The 2005 survey is the first time that a household response rate has been reported because in earlier cycles the adult interview had to be completed before the child or the adolescent interview (i.e., the household rate equaled the adult rate). The adult extended completion rate for 2005 was 54.0 percent, resulting in an overall adult response rate of 26.9 percent for adults. All of the household and person level response rates vary by sampling stratum. For more information about the CHIS 2005 response rates, please see *CHIS 2005 Methodology Series: Report 4 – Response Rates*.

The CHIS response rate is comparable to response rates of other scientific telephone surveys in California, such as the 2005 California Behavioral Risk Factor Surveillance System (BRFSS) Survey with an overall response rate of 29.2 percent. California as a whole and the state's urban areas in particular, are among the most difficult parts of the nation in which to conduct telephone interviews. Survey response rates tend to be lower in California than nationally, and over the past decade response rates have been declining both nationally and in California.

One way to judge the representativeness of a population survey is to "benchmark" its results against those of other reliable data sources. The CHIS 2001 and 2003 samples yielded unweighted and weighted population distributions and rates that are comparable to those obtained from other sources. The demographic characteristics of the CHIS 2001 sample (such as race, ethnicity, and income) were very similar to those obtained from 2000 Census data. CHIS 2001 respondents also had health characteristics and behaviors that were very similar to those found in other extensively used surveys, such as the California BRFSS. The UCLA Center for Health Policy Research is conducting an extensive benchmarking project for CHIS data.

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

Proxy interviews were allowed for frail and ill persons over the age of 65 who were unable to complete the extended adult interview in order to avoid biases for health estimates of elderly persons that might otherwise result. Eligible selected persons were recontacted and offered a proxy option. For 139 elderly adults, a proxy interview was completed by either a spouse/partner or adult child. A reduced questionnaire, with questions identified as appropriate for a proxy respondent, was administered. (Note: questions not administered in proxy interviews are given a value of "-2" in the data files.)

1.5 Weighting the Sample

To produce population estimates from the CHIS data, weights are applied to the sample data

to compensate for the probability of selection and a variety of other factors, some directly resulting from the design and administration of the survey. The sample is weighted to represent the noninstitutionalized population for each sampling stratum and statewide. The weighting procedure used for CHIS 2005 accomplishes the following objectives:

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

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

Population control totals of the number of persons by age, race, and sex at the stratum level for CHIS 2005 were created primarily from the California Department of Finance’s 2004 Population Estimates and 2005 Provisional Population Estimates. The raking procedure used 11 raking dimensions, which are combinations of demographic variables (age, sex, race, and ethnicity), geographic variables (county, Service Planning Area in Los Angeles County, and Health Region in San Diego County), household composition (presence of children and adolescents in the household), and socio-economic variables (home ownership and education). The socio-economic variables are included to reduce biases associated with excluding households without landline telephones from the sample frame. One limitation of using Department of Finance data is that it includes about 2.4 percent of the population of California who live in “group quarters” (i.e., persons living with nine or more unrelated persons). These persons were excluded from the CHIS sample and as a result, the number of persons living in group quarters was estimated and removed from the Department of Finance control totals prior to raking.

1.6 Imputation Methods

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

Two different imputation procedures were used by Westat to fill in missing responses for items essential for weighting the data. The first imputation technique was a completely random selection from the observed distribution of respondents. This method was used only for a few variables when the percentage of the items missing was very small. The second technique was hot deck imputation without replacement. The hot deck approach is probably the most commonly used method for assigning values for missing responses. With a hot deck, a value reported by a respondent for a particular item is assigned or donated to a “similar” person who did not respond to that item. The characteristics defining “similar” vary for different variables. To carry out hot deck imputation, the respondents to an item form a pool of donors, while the nonrespondents are a group of recipients. A recipient is matched to the subset pool of donors based on household and individual characteristics. A value for the recipient is then randomly imputed from one of the donors in the pool. Once a donor is used, it is removed from the pool of donors for that variable. Hot deck imputation was used to impute the same items in CHIS 2003 and CHIS 2005 (i.e., race, ethnicity, home ownership, and education).

The UCLA Center for Health Policy Research imputed missing values for nearly every variable in the data files other than those handled by Westat and some sensitive variables in which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2005 were modest, with most variables missing valid responses for less than 2 percent of the sample. However, there were a few exceptions where item nonresponse rate was greater than 20 percent, such as household income.

The imputation process conducted by UCLA started with data editing, sometimes referred to as logical or relational imputation: for any missing value, a valid replacement value was sought based on known values of other variables of the same respondent or other sample(s) from the same household. For the remaining missing values, hierarchical sequential hot-deck imputation with donor replacement was used. This method replaces a missing value for one respondent using a valid response from another respondent with similar characteristics as defined by a set of control variables. The control variables were

ranked in order from the most to the least important. This procedure allowed control variables to be dropped if certain conditions (such as the minimum number of donors) were not met. The control variables were dropped sequentially, starting from the variable ranked least important. Once a responding case was used as a donor, it was dropped from the donor pool preventing using one donor multiple times.

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

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

1.7 Methodology Report Series

A series of five methodology reports is available with more detail about the methods used in CHIS 2005:

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

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

2. TELEPHONE SAMPLING METHODS

This chapter describes the two general sampling methods used in the CHIS 2005 telephone survey. CHIS 2005 consisted of a telephone random digit dialing (RDD) sample¹ combined with Korean and Vietnamese surname list samples. The RDD sample was drawn using a list-assisted RDD approach, whereas the list samples were drawn from separate surname lists of telephone numbers. The first section describes the list-assisted RDD sampling and the procedures implemented in CHIS 2005 to save costs by reducing the number of calls to ineligible telephone numbers in this sample. The methods implemented were the use of tritone and business purges of unproductive numbers and subsampling of refusals to the screening interview for refusal conversion. The second section describes two noncoverage problems that affect telephone surveys in general and how these were addressed in CHIS 2005. The first is the increasing noncoverage of households with no landline telephone because of cellular telephone use. These households are generally excluded in telephone surveys. The second source of noncoverage is the result of sampling persons who cannot be interviewed because of language limitations. The last section reviews the supplemental samples in CHIS 2005. Lists of surnames were used to supplement the RDD sample to increase the sample size for Koreans and Vietnamese. The RDD was also supplemented with special “child” samples used to increase the number of child interviews.

2.1 List-Assisted Random-Digit-Dial Sampling

List-assisted RDD sampling is currently the standard method of choice for telephone surveys and has been the primary sampling method for each cycle of CHIS. This method results in an unclustered sample that has good operational features (Tucker, Lepkowski, and Piekarski, 2002). In list-assisted sampling, the set of all telephone numbers in operating telephone prefixes is composed of 100-banks, each containing the 100 telephone numbers with the same first eight digits. All 100-banks with at least one residential number listed in a published telephone directory are used to create the sampling frame. A simple random or a systematic sample of telephone numbers is selected from this frame. One disadvantage is a small amount of noncoverage because telephone numbers in 100-banks with no listed telephone numbers are not sampled. Brick et al. (1995) showed that the bias from this approach is considered negligible for most estimates.

¹ Supplemental samples selected by taking proportionally larger samples in certain geographic areas are considered part of the RDD Sample.

When using a list-assisted approach, special data collection procedures are often implemented before data collection to reduce costs and to increase efficiency. One technique implemented in all cycles of CHIS is the use of tritone and business purges of unproductive numbers (i.e., business and nonworking numbers). The procedure used in CHIS 2005, called Comprehensive Screening Service (CSS), is offered by Market Systems Group² (MSG), which also provided the sampling frames.

The CSS process is an enhancement to the procedure called Genesys ID Plus used in CHIS 2003. The performance of this method exceeds that obtained in previous methods. The method classifies a larger proportion of numbers as nonproductive because all telephone numbers, including those identified as listed in the White Pages, are included in the purge. The method also identifies cellular telephone numbers that are excluded in CHIS.

Matches to White and Yellow Pages are used to identify nonresidential business numbers. In the tritone test telephone numbers are dialed to identify nonworking numbers. Any telephone number classified as a nonresidential business number or where a tritone (the distinctive three-bell sound heard when dialing a nonworking number) is encountered in two separate tests is considered nonworking and is not dialed during data collection. Table 2-1 shows the CSS result codes as well as the distribution of the sampled telephone numbers in CHIS 2005. Approximately 45 percent of the sampled numbers (result codes LB, FM, NR NW, and some UB) were excluded from dialing.

Table 2-1. CSSR result codes and their distribution in the CHIS 2005 sample

CSSR result code	Description	Number of telephones	Percentage
CP	Agent dispositioned cell phone	80	0.01
DK	Undetermined	183,028	33.00
FM	Fax/modem	23,175	4.18
LA	Language barrier	3,114	0.56
LB	Listed business	25,647	4.62
NR	No-ring back	2,951	0.53
NW	Nonworking	169,774	30.61
PM	Privacy manager	7,334	1.32
RS	Residence	28,377	5.12
UB	Unlisted business	35,790	6.45
UR	Unlisted residence	74,637	13.46
WR	Pro-t-s wireless detection	665	0.12
Total		554,572	100.00

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

² Market Systems Group (MSG)/Genesys Sampling Systems provides a wide variety of services to the survey research community. Among these services, MSG maintains databases for sample selection in telephone surveys.

The second technique used to reduce costs while improving the sample efficiency in CHIS 2005 was subsampling of refusals for refusal conversion (Brick et al., 2005). In this procedure, a larger sample of telephone numbers than would otherwise be selected is drawn in the first phase. Each number in the first-phase sample is randomly assigned to one of two conditions in the second-phase: subsampled for refusal conversion or not subsampled for refusal conversion. When refusals are encountered at the screening stage of data collection, only numbers in the subsample are eligible for refusal conversion follow-up (at the screener level). The numbers subsampled for refusal follow-up are generally fielded first so that refusal cases can be worked completely (i.e., all of the appropriate scheduling procedures including holding periods for refusal cases can be fully implemented).

The rationale for refusal subsampling depends on two observations: refusal cases comprise the majority of screener nonresponse in CHIS; and substantial effort is expended to gain cooperation in households in which a member refuses to participate in the study at the screener level. The cost savings results from the shift of resources from the less productive labor-intensive task of refusal conversion to the more productive task of completing extended interviews. The principles for refusal subsampling are well established (Hansen and Hurwitz 1946; Elliott, Little, and Lewitzky 2000) and the method has been used in other surveys such as CHIS 2003 and the American Community Survey.

One disadvantage of refusal subsampling is that a weighting adjustment is needed to account for the subsampling, so that those cases that refuse and are subsampled are weighted to represent themselves and the cases that refuse and are not subsampled. This weighting decreases the precision of the survey estimates, but only very slightly. The weighting adjustment is discussed in *CHIS 2005 Methodology Series: Report 5 – Weighting and Variance Estimation*. A subsampling rate of approximately 60 percent was used in CHIS 2005, meaning that 60 percent of the refusal cases at the screener level were eligible for refusal conversion efforts. This subsampling rate of refusal cases is expected to increase the standard error of the estimates less than 3 percent.

2.2 Noncoverage Issues in Telephone Surveys

As in most RDD surveys, households with no access to landline telephones, including those in households with only cellular telephones and households with no telephone service, were not sampled for CHIS 2005³. For estimates correlated with socioeconomic measures such as health insurance

³ A small cellular telephone survey was pilot tested, but not part of the CHIS 2005 study and is not discussed here.

coverage, food security, and poverty, this coverage loss could introduce biases. The bias is related to the percentage of households with no landline telephones and the difference in characteristics of persons in households with and without a landline telephone. The proportion of nonlandline households increased in the recent years due to an increase in households with only cellular telephones (see Blumberg et al., 2006). By 2005, about 10 percent of households had only cellular telephones. This increasing undercoverage presents challenges for future cycles. To reduce potential biases that result from the exclusion of households with no landline telephones, special weighting procedures were used in CHIS 2005 and previous cycles of CHIS. These weighting adjustments are described in *CHIS 2005 Methodology Series: Report 5 – Weighting and Variance Estimation*.

Another source of coverage error in telephone surveys arises when persons who do not speak English are sampled but are not interviewed because of language limitations. These cases are treated as nonrespondents, but could easily be thought of as a coverage problem since none of the persons speaking languages other than those included in the survey protocol are interviewed. In CHIS 2005 and previous cycles, significant efforts have been made to limit this potential bias by interviewing in multiple languages (Lee et al., 2006). In CHIS 2005, interviews were conducted in five languages: English, Spanish, Chinese (Cantonese and Mandarin dialects), Korean, and Vietnamese. This effort should eliminate a large source of the bias that might result from conducting interviews in English only.

2.3 Supplemental Sampling

As mentioned in Chapter 1, one of the goals of CHIS 2005 and previous cycles was to produce reliable estimates for Koreans and Vietnamese in California. These two ethnic groups are important for analytical reasons, but constitute a small proportion of the total California population. As in previous cycles of CHIS, the expected sample yield from the RDD sample was too small to support inferences for these subgroups at the desired level of precision. As in CHIS 2003, two sampling strategies were adopted to meet a target sample yield of 500 for Korean and 500 Vietnamese adult interviews in CHIS 2005: disproportionate stratified sampling and multiple frame sampling (Kalton and Anderson, 1986; and Sudman, Sirken, and Cowan, 1988). Refer to the *CHIS 2003 Methodology Series: Report 1 - Sample Design* for a discussion of other strategies considered, along with the reasons for choosing those that were adopted.

Disproportionate stratified sampling was adopted in 2003 and 2005 for the Korean and Vietnamese samples. Under this scheme, auxiliary information was used to classify telephone exchanges

by the proportion of members of the target groups residing in these exchanges. After classifying the exchanges into strata, the telephone numbers in the exchanges with a relatively high proportion of members (high density stratum) are sampled at a higher rate than the numbers in the other stratum (low density stratum) within selected counties. Refer to the *CHIS 2003 Methodology Series: Report 1 - Sample Design* for additional details on the creation of the substrata.

Disproportionate stratified sampling increased the sample yield for the Koreans and Vietnamese in the base RDD sample; however, this sample was not large enough to meet the goals for these groups. In order to meet the targets, samples from other frames (i.e., surname lists of the race-ethnic groups) have been used since CHIS 2001. This sampling strategy is based on the concept of a dual frame design. In this approach the RDD sample is supplemented with a much less expensive sample from a list of telephone numbers likely to include members of the target group(s). The list frame does not have to be complete to be useful, although the more complete the list the greater the potential for increasing the precision of the estimates. The composition of the list affects its efficiency (that is, the proportion of sampled numbers that lead to a member of the target group), but not the ability to produce unbiased estimates. Unbiased estimates can be produced if the list membership of every sampled unit (telephone number) from the other frame (RDD in our case) can be determined. The cost associated with the use of the surname lists was much lower than the cost for locating and interviewing members of the groups from the base RDD sample.

The identification of eligible (i.e. Korean or Vietnamese) adults in the list samples is done through questions in the screener interview. This strategy is relatively simple to implement and has good statistical properties, except for a measurement error that may be introduced by asking a question about the ethnicity of the adults at the beginning of a telephone interview. Screening is not necessary for the cases sampled from the high/low density strata because these cases are part of the base RDD sample where all households are eligible for further interviewing. Although the use of surname lists was an effective way to increase the number of completed cases for these groups, the variances of the estimates for these groups is not greatly reduced by this approach. Refer to the *CHIS 2003 Methodology Series: Report 1 - Sample Design* for additional details on the use of the list samples.

Supplemental sampling was also used in CHIS 2005 to increase the number of child interviews in the state and in San Diego County. These samples are referred to as “child supplemental” samples. As in the surname supplemental samples, questions in the screener interview are used to identify households with children. One disadvantage of this approach is the presence of noncoverage error from households that misreport the presence of children. Another disadvantage is that if the proportion of

households with children is small, then a large number of households need to be screened in order to achieve the goals of the study.

3. SAMPLING HOUSEHOLDS

This chapter describes the sample design and selection of households for CHIS 2005. We begin by defining the target population and the persons included and excluded in the survey. Target numbers of completed adult interviews by county and for the supplemental samples are then described. The remainder of the chapter describes the types of supplemental samples and how the telephone numbers were selected in order to achieve the stated goals. The last section reviews the statistical issues considered in arriving at the allocation of the sample by county.

3.1 Population of Interest

As in previous CHIS cycles, the 2005 sample was intended to represent the adult (age 18 and older) residential population of California, as well as adolescents (age 12-17) and children (age 11 and under). Eligible residential households included houses, apartments, and mobile homes occupied by individuals, families, multiple families, extended families or multiple unrelated persons, provided that the number of unrelated persons was less than nine. Persons living temporarily away from home were eligible and enumerated at their usual residences. These include college students in dormitories, patients in hospitals, vacationers, business travelers, and so on. The survey excluded group quarters – any unit occupied by nine or more unrelated persons (e.g., communes, convents, shelters, halfway houses, or dormitories). Institutionalized persons (e.g., those living in prisons, jails, juvenile detention facilities, psychiatric hospitals and residential treatment programs, and nursing homes for the disabled and aged), the homeless, persons in transient or temporary arrangements, and those in military barracks were also excluded. As described in Chapter 2, some individuals who were part of the residential population did not have a chance of selection, including those living in households without landline telephones (either without any telephone service or with cellular telephone service only), and children and adolescents living in a household without a parent or legal guardian.

3.2 Sample Design

The principal goals of the CHIS 2005 sample design were (1) to produce reliable statewide estimates for the total population in California and for its larger race/ethnic groups, as well as for several

smaller ethnic groups (i.e., Koreans and Vietnamese), and (2) to produce reliable estimates at the county level for as many counties as possible. In CHIS 2005 and also for the previous cycles in CHIS, a base RDD sample and several supplemental samples were drawn in order to meet these goals. The supplemental samples included in CHIS 2005 were the surname list samples, geographic samples and child samples. Details of the base RDD and supplemental samples are described in the following sections.

3.2.1 Base RDD Sample

The CHIS 2005 base RDD sample had an initial goal of completing 37,380 adult interviews statewide. When more funding became available, the goal was augmented by 1,703 interviews in the 14 largest strata to a total of 39,083. Because in previous CHIS cycles it had proven difficult to control the data collection closely enough to meet the stratum goals exactly, an adjusted goal of 97.5 percent of the original goal was set in most strata; the exceptions were Los Angeles County and strata with supplemental geographic samples. Thus, the overall adjusted goal was 38,444 adult interviews, with the expectation that the adjusted goals would be exceeded in some strata. The base RDD sample goal for adult interviews in CHIS 2005 was approximately 1,250 interviews higher than the base RDD sample goal for CHIS 2003.⁴ For the first time in the CHIS surveys, there was a target for the number of child interviews in 2005. The revised target was 8,326 completed child interviews conducted with knowledgeable parents or guardians. Although the number of adolescent interviews was not predetermined, 3,000 to 4,000 completed interviews were expected (depending on compliance since parental consent and adolescent agreement are required) from the base RDD sample.

The goals of the base RDD sample required a compromise in the sample allocation into the sampling strata. To achieve the most reliable statewide estimates, the optimal design is to allocate the sample to counties proportional to their population. On the other hand, the optimal allocation for producing county-level estimates is to assign each county an equal sample size.

The stratification of California's 58 counties used in CHIS 2005 was different from that used in previous CHIS cycles. The geographic stratification was altered to increase the number of independent counties that formed their own sampling stratum. Eight counties that had previously been combined in multi-county strata were split to become single-county strata, increasing the total number of strata from

⁴ Goals excluded the geographic samples in CHIS 2003 and 2005.

41 to 44. The new sample design includes 41 single-county strata (compared to 33 in the original design) with the remaining counties grouped into three geographic strata that combined two or more counties (compared to 8 multi-county strata in the original design). Table 3-1 shows the 44 geographic sampling strata, the original and adjusted target numbers of adult and child interviews for CHIS 2005. Table A-1 in the appendix shows the assignment of counties to geographic strata across the CHIS cycles.

Because of the need to produce reliable estimates at the county level, the sample allocation is not in all cases proportional to the population across counties. With a proportional allocation, the estimates from the smaller counties would be based on small sample sizes and would not be adequate for the envisioned analyses. To achieve the goal of producing local or county estimates, the sample sizes from the largest counties are re-distributed to the smaller counties. The adjusted target sample sizes ranged from 9,000 adult interviews in Los Angeles to 390 interviews in the smallest strata.

Table 3-1. Targeted number of complete adult and child interviews for the base RDD sample by county

Stratum	Targeted number of interviews			Population size
	Adult		Child	
	Original	Adjusted	Adjusted	
1 Los Angeles*	9,000	9,000	1,833	Over 9,000,000
2 San Diego	2,622	2,622	548	
3 Orange	2,396	2,336	517	
4 Santa Clara	1,415	1,380	305	
5 San Bernardino	1,420	1,385	341	1,200,000 or greater
6 Riverside	1,415	1,380	329	
7 Alameda	1,321	1,288	276	
8 Sacramento	1,209	1,179	238	
9 Contra Costa	884	862	181	800,000 to 1,200,000
10 Fresno	663	646	196	
11 San Francisco	820	800	107	
12 Ventura	628	612	134	
13 San Mateo	623	607	117	500,000 to 800,000
14 Kern	567	553	135	
15 San Joaquin	500	488	119	
16 Sonoma	500	488	102	
17 Stanislaus	500	488	118	Medium counties 100,000 to 500,000
18 Santa Barbara	500	488	115	
19 Solano	500	500	113	
20 Tulare	500	488	135	
21 Santa Cruz	500	488	92	
22 Marin	500	500	101	

Table 3-1. Targeted number of complete adult and child interviews for the RDD sample by county (Continued)

Strata	Targeted number of interviews			Population size
	Adult		Child	
	Original	Adjusted	Adjusted	
23 San Luis Obispo	500	488	89	Medium counties 100,000 to 500,000
24 Placer	500	488	104	
25 Merced	500	488	148	
26 Butte	500	488	94	
27 Shasta	500	488	87	
28 Yolo	500	488	107	Small counties less than 100,000 population per county
29 El Dorado	500	488	100	
30 Imperial	500	488	129	
31 Napa	500	488	93	
32 Kings	500	488	164	
33 Madera	500	488	110	
34 Monterey	500	488	126	
35 Humboldt	500	500	94	
36 Nevada	400	390	49	
37 Mendocino	400	390	69	
38 Sutter	400	390	97	Small counties combined
39 Yuba	400	390	102	
40 Lake	400	390	83	
41 San Benito	400	390	109	
42 Colusa, Glenn, Tehama, Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou,	400	390	93	
43 Trinity Amador, Alpine, Calaveras, Inyo, Mariposa, Mono,	400	390	71	
44 Tuolumne	400	390	56	
Total	39,083	38,444	8,326	

*The Los Angeles County sample includes the 250 additional cases in Antelope Valley.

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

3.2.2 Supplemental Geographic Samples

In CHIS 2005 supplemental geographic samples were added at the request of Solano

County, Marin County and Humboldt County after funding was arranged. Officials in these counties were interested in larger samples for a more detailed analysis. Since the geographic samples covered entire counties, they were deemed as part of the base RDD sample for the specific county and were drawn in the same way as the base RDD sample. Unlike some supplemental geographic samples in previous cycles, screening was not used to determine if the case was in the county of interest. As a result, there was no difference between the base RDD sample and supplemental geographic sample instruments.

Table 3-2 shows the targeted number of adult and child interviews for the supplemental geographic samples in CHIS 2005. Combining the base RDD and geographic samples the total targeted number of adults was 1,230 for Solano County, 3,000 for Marin County, and 800 for Humboldt County.

Table 3-2. Targeted number of interviews for the supplemental geographic samples

Stratum	Targeted number of interviews	
	Adult	Child
19 Solano	730	165
22 Marin	2,500	505
35 Humboldt	300	56
Total	3,530	726

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

3.2.3 Supplemental Surname List Samples

The second type of supplemental sample includes the list surname supplemental sample used to increase the number of completed interviews of adults of Korean and Vietnamese descent. These were the only two race-ethnic groups with statewide sample goals in CHIS 2003 and 2005 not expected to be achieved from the base RDD sample, with a goal of 500 completed adult interviews from each group. To achieve these goals, the base RDD sample was supplemented with samples drawn from lists containing Korean and Vietnamese surnames. Screening was used in the surname samples to determine if there were eligible adults (i.e., adults of Korean or Vietnamese descent) in the household. If there were no eligible adults, the interview was terminated and the case was coded as ineligible.

Table 3-3 shows the sampling goals for completed adult interviews for Koreans and Vietnamese in CHIS 2005. The surname list sample targets were adjusted during data collection as the actual RDD yield became known.

Table 3-3. Targeted number of complete adult interviews for the Korean and Vietnamese supplemental list samples

Subgroup	Targeted number of adult interviews		Total
	Base RDD sample	Supplemental list sample	
Korean	377	123	500
Vietnamese	373	127	500
Total	750	250	1,000

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

The sampling list frames were created by the sampling vendor MSG by compiling lists of surnames likely to be Korean or Vietnamese from telephone directories in California. MSG provided three nonoverlapping surname frames; the first two frames included telephone numbers whose associated surnames were very likely to be Korean (frame 1) or Vietnamese (frame 2). The third frame included those surnames likely to be either Korean or Vietnamese. Separate samples were drawn from each of the three frames. Households were deemed eligible if they included an adult who was either Korean or Vietnamese, regardless of which frame the number was drawn from. Table 3-4 shows the size of the surname list frames used in 2005.

Table 3-4. Number of records in the surname frames

Surname frame	Number of records
Korean only	137,149
Vietnamese only	102,740
Korean or Vietnamese	56,313

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

3.2.4 Supplemental Child Samples

The third type of supplemental sample in CHIS 2005 was drawn to increase the number of child interviews. Two child supplemental samples were implemented in CHIS 2005: a statewide oversample sponsored by First 5 California and another in San Diego County sponsored by the San Diego County Health and Human Services Agency. Screening these supplemental samples for the presence of children and retaining only those households with children eliminated unnecessary interviews of adults in households with no children. Once the household was retained, the same selection procedures used in the

base RDD sample were used to select the adult, child⁵ and adolescent (if present) in the household. In contrast, if the household was part of the base RDD sample, then an adult was always selected regardless of presence of children in the household. Table 3-5 shows the sample goals for the number of adult and child interviews for the child supplemental samples in 2005.

Table 3-5. Targeted number of complete adult and child interviews for child supplemental samples

Child supplemental samples	Targeted number of interviews	
	Adult	Child
State	562	562
San Diego	1,166	1,166

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Table 3-6 shows the details of the total target number of adult and child interviews for all the RDD samples in San Diego.

Table 3-6. Targeted number of complete adult and child interviews for San Diego County

Samples	Targeted number of interviews	
	Adult	Child
Base RDD sample	2,622	548
Child supplemental sample		
State (expected sample in San Diego County)	69	69
San Diego	1,435	1,435
Total	4,126	2,052

Table 3-7 shows the total sample goals for the number of adult and child interviews for all the samples (i.e., base RDD sample, geographic, surname list, and child samples) in CHIS 2005.

⁵ Note that all households with children were eligible for the supplemental samples, even if the adult selected had no associated children in the household. Thus, for some households in the child supplemental samples an adult interview was completed but no child was sampled.

Table 3-7. Targeted number of complete adult and child interviews for all samples

Samples	Targeted number of interviews	
	Adult	Child
Base RDD sample	38,444	8,326
Geographic samples (Solano, Humboldt, and Marin)	3,530	726
Surname list samples (Korean and Vietnamese)	250	NA
Child supplemental sample (State and San Diego)	1,728	1,728
Total	43,952	10,780

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

At the beginning of the study, different allocations of the sample consistent with the available budget were evaluated. The UCLA CHIS staff consulted with various constituencies to assess the relative importance of particular types of estimates. Westat statistical staff helped evaluate each alternative and examined the consequences of the sample allocations. The main statistical issues were communicated by computing effective sample sizes for the main groups for the alternative designs. The expected effective sample size computations are discussed in Section 3.5.

3.3 Stratification

In this section, we describe the steps used in selecting the sample of telephone numbers for CHIS 2005. These steps include stratifying the telephone numbers by sampling stratum, selecting the sample of numbers after adjusting for expected losses due to nonresponse, and subsampling the numbers based on mailable status and refusal status to improve the efficiency of the sample.

Since CHIS 2005 used a stratified sample, the first step was stratifying the sampling frame of 100-banks with one or more listed telephone numbers into nonoverlapping strata, each corresponding to a county or a group of counties as shown in Table 3-1. The same procedure was used in previous cycles of CHIS. The geographic information required for stratification is available only at the exchange level⁶, so 100-banks could not be assigned directly to a single stratum. All banks within an exchange were stratified indirectly by mapping the exchanges to a county represented by the stratum. However, some telephone exchanges actually service households in more than one county.

⁶A telephone exchange consists of 10,000 consecutive telephone numbers with the same first six digits including area code. An exchange is a set of area codes and prefixes serving the same geographic area.

To solve the stratification problem, the procedure used coverage reports for each county in California produced by Genesys. The coverage reports listed all the exchanges in the county. For each exchange, the report gave the total number of listed households in the exchange and the proportion of listed households that are within the county. After combining the information of the coverage reports for all 58 counties, we created a frame of exchanges with variables for the number of listed households in each county that the exchange covers. Each exchange was assigned to the county with the most listed households. As in 2003, there was also interest in obtaining a better sample distribution for Los Angeles County by Service Planning Areas (SPAs) in 2005. Using ZIP Code information, telephone exchanges in Los Angeles were classified into eight subsampling strata, each representing a SPA. Telephone exchanges that crossed SPAs were assigned to the SPA with the most listed households. There were no targets for individual SPAs, so the sample for Los Angeles was allocated proportionally by these substrata, except for the sample for Antelope Valley. The sample for Antelope Valley included an additional sample to yield 250 adult interviews more than what would be expected from the proportional allocation.

As mentioned in Chapter 2, disproportionate stratified sampling was used to oversample Koreans and Vietnamese without increasing the sample size allocated to any stratum (the stratum sample size was fixed). The CHIS 2003 approach to oversample geographic areas with high concentrations of Korean and Vietnamese was modified slightly for CHIS 2005. In this approach high and low concentration substrata were created in the four sampling strata (counties) where the Korean and/or Vietnamese population was large enough to produce increases in the expected number of interviews. The sampling substrata covered approximately 78 percent of the Korean and Vietnamese population in California while the oversampled exchanges represented less than 40 percent of the Korean and Vietnamese population. The analysis done in CHIS 2003 found that six percent or more Korean or Vietnamese in the exchanges was optimal for the creation of the substrata. In addition, the analysis showed that oversampling the substrata with high concentration at twice the rate of the low concentration strata did not inordinately inflate the design effect nor decrease the effective sample sizes for other race-ethnic groups of interest that were not oversampled. See *CHIS 2003 Methodology Series: Report 1 - Sample Design* for additional details of the analysis for the creation of high and low density substrata.

Since the creation of the high/low density designation used information from Census 2000, the assignment of telephone exchanges was revised in CHIS 2005. Tabulations of the number of Korean or Vietnamese interviews by telephone exchange were produced using data from previous CHIS cycles. Using this information, some exchanges were reallocated to the high/low density strata depending on the number of interviews completed with adults of Korean or Vietnamese descent. The high/low density

subsampling strata were created in San Diego County, Orange County, and Santa Clara County. Fourteen substrata were created in Los Angeles County by classifying the SPAs into high/low density substrata.

One month after the beginning of the data collection, the target sample size for child interviews in San Diego County was increased. The sample design required the number of child interviews from the RDD and child supplemental samples combined to be approximately of the same size in each of the six San Diego Health and Human Services Agency (HHSA) Service Regions. Because the substrata were already created using the high/low density areas, we proceeded to divide them by HHS Service Region creating eight new substrata in the county. The sample selection for the additional cases had to consider the fact that telephone numbers from the base RDD sample were already selected and fielded in the county. Additional telephone numbers in the eight substrata in San Diego County were released sequentially depending on the number of completed child interviews that had been achieved during data collection. Table 3-8 shows the definition of the substrata for Los Angeles County, San Diego County, Orange County and Santa Clara County. The table also shows the number of telephone exchanges and the estimated number of households in the substrata.

Table 3-8. Definition of sampling substrata, number of exchanges, and total number of households for Los Angeles County, San Diego County, Orange County, and Santa Clara County

Stratum	Substratum	SPA/Service Region	Density	Number of telephone exchanges	Number of households
1. Los Angeles	1.012	San Fernando SPA	High	30	54,265
	1.013	San Gabriel SPA	High	59	114,804
	1.014	Metro SPA	High	76	102,937
	1.017	South SPA	High	21	23,195
	1.018	South Bay SPA	High	37	47,774
	1.021	Antelope Valley SPA	Low	37	69,921
	1.022	San Fernando SPA	Low	327	492,580
	1.023	San Gabriel SPA	Low	196	298,550
	1.024	Metro SPA	Low	171	223,188
	1.025	West SPA	Low	200	206,292
	1.026	South SPA	Low	149	267,540
	1.027	East SPA	Low	140	248,629
	1.028	South Bay SPA	Low	214	316,525
2. San Diego	2.012	North Central Service Region	High	22	28,238
	2.013	Central Service Region	High	11	17,946
	2.021	North Coastal Service Region	Low	85	144,878
	2.022	North Central Service Region	Low	109	118,304
	2.023	Central Service Region	Low	74	121,679

Table 3-8. Definition of sampling substratata, number of exchanges, and total number of households for Los Angeles County, San Diego County, Orange County, and Santa Clara County (Continued)

Stratum	Substratum	SPA/Service Region	Density	Number of telephone exchanges	Number of households
	2.024	South Service Region	Low	42	87,691
	2.025	East Service Region	Low	59	113,381
	2.026	North Inland Service Region	Low	79	142,367
3. Orange	3.01		High	188	233,339
	3.02		Low	362	538,298
4. Santa Clara	4.01		High	175	220,679
	4.02		Low	253	262,263
Total				3,116	4,495,263

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

3.4 Sample Selection

The number of telephone numbers selected in any RDD survey has to be greater than the targeted number of completed interviews to account for a variety of factors. For example, a substantial percentage of the sampled telephone numbers is not residential. For CHIS 2005 the sample of telephone numbers was inflated to deal with losses due to the following sources:

- Nonworking, nonresidential, and never answered numbers;
- Subsampling for refusal conversion;
- Nonresponse to the screening interview; and
- Nonresponse to the adult extended interview and to the child interview for the child supplemental samples.

The first, third, and fourth sources noted above are typical of all RDD surveys. To deal with these losses we used information from CHIS 2003 to estimate the percentage of the telephone numbers that would not be residential and the percentage that would not respond to the screener and extended interviews, and increased the sample size accordingly.

The only source of loss that requires additional discussion is the subsampling for refusal conversion. As mentioned in Chapter 2, during CHIS 2005 sample selection 60 percent of the telephone

numbers were flagged for refusal conversion. Refusal conversion efforts were made only to flagged telephone numbers after the respondent refused to do the screener interview. Taking all of these factors into consideration, a total of 554,572 telephone numbers⁷ were drawn and dialed in CHIS 2005. Not all the telephone numbers were selected at the same time as the sample design was modified several times during the field period. After each selection, duplicate telephone numbers were removed from the samples. Telephone numbers from the base RDD sample also selected in other samples were regarded as part of the base sample. The data collection procedures are discussed in *CHIS 2005 Methodology Series: Report 2 - Data Collection Methods*. Table 3-9 summarizes the size of each type of sample. Table A-2 in The appendix shows the sample size by sampling stratum for the different samples.

Table 3-9. Number of telephone numbers drawn by type of sample

Sample type	Number of telephone numbers drawn
Base RDD sample	432,446
Geographic supplemental samples	
Solano	700
Marin	32,575
Humboldt	1,900
Surname List samples	
Korean list	1,832
Vietnamese list	1,438
Korean/Vietnamese list	1,608
Child supplemental samples	
State	23,787
San Diego	58,286
Total	554,572

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

3.5 Expected Design Effect

Sections 3.2 and 3.3 described the allocation of the sample of telephone numbers by sampling stratum and substratum and noted that it was a compromise among three goals: to produce reliable estimates for the entire state, to produce estimates at the county level, and to oversample Koreans and Vietnamese. Allocating the sample proportionally to the population in the counties would be approximately optimal for statewide estimates. For county estimates, an equal allocation would be more efficient. In this section, we describe the statistical methods used to examine the efficiency of the sample

⁷ This total includes all samples in CHIS 2005.

under different allocations. These methods helped guide the allocation of the CHIS 2005 sample.

If CHIS 2005 had been a simple random sample, it would be relatively simple to predict the precision of the estimates. Under the assumption of simple random sampling, suppose we wish to estimate a proportion of adults with a characteristic, say p . If the sample size is large enough, then the standard $(1-\alpha)\cdot 100$ percent confidence interval of the estimated proportion, \hat{p} , is

$$\left(\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) \quad (1)$$

where $z_{1-\alpha/2}$ is the critical value from the standard normal distribution, and n is the number of completed interviews. This form of the confidence interval is not appropriate for CHIS 2005 for several reasons. The main reason is that the allocation of the sample to the counties does not produce a simple random sample across the state. Other reasons why the estimated proportion given in (1) is not fully appropriate are sampling within households and other adjustments made to the weights. These issues are covered in *CHIS 2005 Methodology Series: Report 5 - Weighting and Variance Estimation*.

To adjust (1) to account for the sample allocation to the counties or strata we introduce the concept of a design effect. Kish (1992) discusses the design effect in some detail. Here we simply note that in stratified designs like CHIS, the design effect measures the departures with respect to a sample proportionally allocated across the strata. A sample with proportionate allocation has a design effect of one. Departures from proportionate allocation result in design effects greater than one.

The design effect due to departures from proportional allocation can be computed as

$$D = \left(\sum_{h=1}^H W_h k_h \right) \left(\sum_{h=1}^H \frac{W_h}{k_h} \right), \quad (2)$$

where W_h is the proportion of the population in sampling stratum h computed as $W_h = N_h (\sum N_h)^{-1}$, where N_h is the population total in stratum h , and k_h is the relative sampling rate for stratum h . More specifically, k_h is defined as $k_h = \frac{n_h}{N_h} \frac{N_1}{n_1}$, where n_h is the sample size in stratum h and the reference stratum is set to be stratum 1 so that $k_1 \equiv 1$ (the choice of the reference stratum does not affect the computations since the relative sampling rates are the only factors involved).

Using the design effect computed this way we can estimate the effective sample size for a stratified sample with a given allocation. The effective sample size is the number of cases needed from the stratified sample to produce estimates with the same precision that would be expected from a simple random sample design. The effective sample size n_{eff} is computed as

$$n_{eff} = \frac{n}{D}. \quad (3)$$

where n is the nominal sample size and D was defined above.

In CHIS 2005, we expected to complete 38,456 adult interviews from the RDD sample (the supplemental geographic samples, supplemental child samples, and the supplemental list samples were not included in this evaluation). The expected nominal sample sizes (the number of adult interviews), the expected design effects due to the sample allocation to the strata using (2), and the expected effective sample sizes using (3) are given in Table 3-10. The expected design effects and effective sample sizes are given for the entire state and for domains defined by race and ethnicity. It is important to remember that the design effects are computed at the household level and do not include any adjustments for nonresponse, within-household sampling, or other weighting adjustments.

Table 3-10. Expected design effects and effective adult sample size associated with the sample allocation for the base RDD sample

Race and ethnicity	Expected nominal sample size	Expected design effect	Expected effective sample size
White alone	25,571	1.23	20,859
African-American alone	2,243	1.08	2,086
American Indian/Alaskan Native alone	538	1.33	405
Asian alone	3,609	1.12	3,216
Native Hawaiian/Pacific Islander alone	98	1.14	86
Other race alone	4,979	1.21	4,124
Two or more races	1,405	1.24	1,137
Latino	9,554	1.19	8,051
All	38,444	1.20	31,940

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

For example, the expected yield from the CHIS 2005 sample for African-Americans was 2,243 adults for the RDD sample. Due to the allocation of the sample, the expected effective sample size was 2,086. The 95 percent confidence interval for an estimated proportion can be computed by using the entries in this table and replacing n in (1) by n_{eff} . For example, for estimating a proportion of $p = 0.5$ for

American Indian/Alaska Natives, the 95 percent confidence interval is

$$\left(0.5 - 1.96\sqrt{\frac{0.5^2}{538}}, 0.5 + 1.96\sqrt{\frac{0.5^2}{538}} \right) = (0.4577, 0.5423)$$

As the UCLA CHIS staff consulted with various groups in California to evaluate the data needs that CHIS could help to support, they developed different allocation schemes for distributing the sample to the counties. The effects of these allocations were examined by using the methods presented above. The UCLA CHIS staff then chose the sample allocation that best satisfied the needs of survey data users.

4. WITHIN-HOUSEHOLD SAMPLING

Once the sample of telephone numbers is selected, interviewers call the numbers and select and conduct interviews with sampled persons within the household. This chapter describes the procedures for selecting the sample of persons within households for CHIS 2005. Samples of adults, children, and adolescents within the household were selected using different sampling procedures, but one adult and up to one child and one adolescent were sampled within each household. The within-household sampling procedures were developed to maximize the analytic utility of the data collected from the respondents.

The next section describes the within-household sampling alternatives we evaluated and the reasons for choosing the specific method of sampling. The second section describes the operational “child-first” procedure used to increase the number of child interviews. The last sections describe the methodology used for sampling adults, children, and adolescents in CHIS 2005.

4.1 Sampling Alternatives

The general idea for the sample design over the CHIS cycles has been to sample randomly one adult from all the adults in each sampled household. In addition, in those households with adolescents (ages 12-17) and/or children (under age 12), one adolescent and one child were to be sampled and interviewed (a parent of the child was interviewed about the child). One approach to accomplishing these goals is to simply list all the persons in the age group (adult, child, and adolescent) in the household and select one person randomly from each group. We call this the *completely random* sampling method.

The completely random sampling method is not a problem in most households because most households have only one family. However, in households with two or more families, the completely random method could result in selecting persons from the different age groups who were not members of the same family. This situation is undesirable because the adult interview collects data about the family of the sampled adult. The data from the adult interview are of great value for the analysis of the data from the child and adolescent interviews. If the sampled child and/or sampled adolescent were not members of the same family as the sampled adult, then the data collected about them would be of very limited utility.

To resolve this analytic problem, a second sampling alternative was adopted in CHIS 2001

and continued in 2003 and 2005. We call this method the *linked* sampling approach. In this approach, children and/or adolescents for whom a sampled adult was a blood or adoptive parent, or a legal guardian were considered as linked to or “associated” with that adult.

In the linked sampling method persons are sampled in two phases. In the first phase, an adult is randomly sampled from all the adults in the household. In the second phase, a child is sampled from all the children associated with the sampled adult. Since the sampling of children is a two-phase procedure, the probability of sampling the child is the product of the probability of sampling the adult (phase one) and the probability of sampling the child from all children associated with that adult (phase two). Adolescents are sampled in the same way, that is, one adolescent is sampled from all adolescents associated with the adult sampled in the first phase.

To use the linked sampling method, data are needed to link children and adolescents in a household to the sampled adult and his/her spouse/partner (children or adolescents linked to both the sampled adult and spouse/partner could be selected if either adult was sampled). These data were collected during the screener interview or the adult interview in CHIS 2005. We expected that in a very few households it would not be possible to link or associate a child or adolescent to an adult because of unusual household structures. A child or adolescent not associated with an adult would not have a chance of being selected. Beginning in 2003, the UCLA Institutional Review Board (IRB) directed that only children and adolescents of the sampled adult could be selected. Therefore, unassociated children and adolescents in a household could not be randomly linked to an adult in the household in 2003 or 2005. The bias due to excluding unassociated children and adolescents was expected to be very small; however, due to changes in the way adults, children and adolescent were enumerated beginning in 2003 cycle, it is not possible to evaluate this bias.

4.2 Child First Procedure

In previous cycles of CHIS, children and adolescents were enumerated and sampled during the adult extended interview. The child and/or adolescent interviews were then conducted following the adult interview. Beginning in 2005, changes in the CATI permitted the child and adolescent interviews to be conducted prior to the adult interview under certain conditions. These changes in the order a child and/or adolescent was selected and interviewed are called the “child-first” procedure. This procedure was an operational method (not a sampling method) used to increase the sample yield for child interviews. In

previous cycles of CHIS children and adolescents were enumerated and sampled at about the mid-point of the adult interview (section G). If the adult did not complete the extended interview, the child and adolescent could not be interviewed. The child-first procedure was used only when the screener respondent was the spouse or partner of the sampled adult, there were children in the household, and the sampled adult was not available at the time of the interview. If these conditions were met, a child and or adolescent could be sampled and the appropriate interview was conducted without waiting for the completion of the adult interview. When the child-first criteria were not met, the adult was interviewed as in previous cycles of CHIS.

Table 4-1 shows the distribution of completed screener interviews for households with children and the number of households where the child-first procedure was used in CHIS 2005. In CHIS 2005, 73,775 households completed the screener interview. The child-first procedure was used in 32 percent of the completed screener interviews with households with children. A child interview was completed in 72.0 percent (5,037 interviews) of the households with children where the child-first procedure was used. In comparison, a child interview was completed in only 38.2 percent (6,321 cases) of the households with children where the procedure was not used.

If we assume that the child-first procedure had not been used and that the completion rate for child-first households would have been the same as for nonchild-first households, the child-first procedure increased the number of completed child interviews by over 2,300, or approximately 26 percent. Although the child-first procedure was intended to increase the number of child interviews, it had also an effect on the number of adolescent interviews. The child-first procedure was used in 13.8 percent of the households with adolescents and 40.3 percent of those completed the adolescent interview. In contrast, in only 24.1 percent of the households with adolescents where the child first procedure was not used the adolescent interview was completed. Again assuming that the completion rate for child-first households with adolescents would have been the same as for other households with adolescents, we estimate that the procedure produced an increase of 340 cases (10%) in the adolescent interviews.

See *CHIS 2005 Methodology Series: Report 2 - Data Collection* for more detail on the child-first procedures and further evaluation of the yields.

Table 4-1. Effect of the child-first procedure on completed child and adolescent interviews

Households that completed a screener interview	Count	Percentage
Total number of households	73,775	100.0
Households with children	23,500	31.9
Households with adolescents	15,307	20.7
<hr/>		
Households with children that completed a screener interview	Count	Percentage
Total number of households with children	23,500	100.0
Participated in the child-first procedure	6,998	29.8
Did not participate in the child-first procedure	16,502	70.2
<hr/>		
Households with children that completed a screener interview and participated in the child-first procedure	Count	Percentage
Total number of households participating in the child-first procedure	6,998	100.0
Completed the extended interview	5,037	72.0
Did not complete the extended interview	1,961	28.0
<hr/>		
Households with children that completed a screener interview and did not participate in the child-first procedure	Count	Percentage
Total number of households not participating in the child-first procedure	16,502	100.0
Completed the extended interview	6,321	38.3
Did not complete the extended interview	10,181	61.7
<hr/>		
Households with adolescents that completed a screener interview	Count	Percentage
Total number of households with adolescents	15,307	100.0
Participated in the child-first procedure	2,106	13.8
Did not participate in the child-first procedure	13,201	86.2
<hr/>		
Households with adolescents that completed a screener interview and participated in the child-first procedure	Count	Percentage
Total number of households participating in the child-first procedure	2,106	100.0
Completed the extended interview	848	40.3
Did not complete the extended interview	1,258	59.7
<hr/>		
Households with adolescents that completed a screener interview and did not participate in the child-first procedure	Count	Percentage
Total number of households not participating in the child-first procedure	13,201	100.0
Completed the extended interview	3,181	24.1
Did not complete the extended interview	10,020	75.9

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

4.3 Adult Sampling

For CHIS, an adult is defined as any person 18 years or older residing in the household. The procedure to select adults in CHIS 2005 was the same as that used in 2003, called the Rizzo method (see Rizzo et al., 2004, for a complete discussion of the method and its implementation). The principal advantage of this method is that the enumeration of adult household members is bypassed in most households, so it is less intrusive while still resulting in a valid probability sample. In this method, all sampled adults have an equal probability of selection. A sampled adult is selected using the following steps:

- Ask the screener respondent (who must be an adult living in the household) how many adults are in the household (i.e., N). The respondent answers $N = 1, 2, 3, \dots$;
- If there is only one adult in the household (i.e., $N = 1$), then that adult is selected;
- If there are two adults in the household (i.e., $N = 2$), then the CATI system accesses a pre-generated uniform random number between 0 and 1.
 - If the random number is less than or equal to 0.5 then the screener respondent is selected;
 - If the random number is greater than 0.5 then the other adult is selected;
- If there are more than two adults in the household (i.e., $N > 2$), then the CATI system accesses a pre-generated uniform random number between 0 and 1.
 - If the random number is less than or equal to $1/N$ (i.e., the inverse of the number of adults in the household) then the screener respondent is selected;
 - If the random number is greater than $1/N$, then the screener respondent is asked which of the other adults is the next to have a birthday; and
 - * If the screener respondent knows which of the other adults is next to have a birthday, then the adult with the next birthday is selected.
 - * If the screener respondent does not know which of the other adults is next to have a birthday then the screener respondent is asked to list the adults in the household (excluding himself/herself) and the CATI system randomly chooses one of the adults from this roster.

If the number of adults in the household is unknown then the screener respondent is asked to list the adults in the household (including the screener respondent) and the CATI system randomly chooses one of the adults from this roster. No other sampling steps are necessary.

4.4 Child Sampling

In previous cycles of CHIS, if there were any children under age 12 in the household who were associated with the sampled adult, then exactly one child was sampled and each associated child had an equal probability of selection. In CHIS 2005 the child sampling procedure was modified in order to increase the number of interviews for younger children (0 to 5 years old) while reducing the number of interviews for older children (6 to 11 years old). If there were only younger or older children in the sampled households, a child was selected with equal probability of selection as in previous cycles of CHIS. In contrast, in households with both younger and older children, children were sampled with differential probabilities of selection. Younger children in such households were assigned a greater probability of selection with respect to the older children. The probability assigned to children i in the household h , p_{hi} , was assigned as

$$p_{hi} = \begin{cases} \frac{2NC_{1h}}{2NC_{1h} + NC_{2h}} & \text{If age of child } i \text{ in household } h \text{ is between 0 and 5 years old (younger child)} \\ \frac{NC_{2h}}{2NC_{1h} + NC_{2h}} & \text{If age of child } i \text{ in household } h \text{ is between 6 and 11 years old (older child)} \end{cases}$$

where NC_{1h} is the number of younger children and NC_{2h} is the number of older children in the household h . For example, in a household with one young child and one older child, the young child was twice as likely to be selected as the older child. The disadvantage of this approach was that the number of interviews about older children was reduced and there was a slight increase in the design effect for estimates for all children due to the disproportionate sampling.

Table 4-2 shows the number of households with a completed screener interview in which the enumeration and selection of children were completed (either at the end of the extended interview for child-first cases or in section G of the adult extended interview) in CHIS 2005. Children were selected with unequal probability of selection in approximately 26 percent of the households with children. Comparing the results of CHIS 2003 and 2005, the proportion of completed interviews for young children among all child interviews increased from 46 percent to 51 percent.

Table 4-2. Distribution of households with children by type of child sampling

Type of child sampling	Type of household	Number of households	Percentage
Equal probability	Household with children 0 to 5 years old	5,416	33.4
	Household with children 6 to 11 years old	6,538	40.3
Unequal probability	Household with children 0 to 5 and 6 to 11 years old	4,249	26.3
Total		16,320	100.0

4.5 Adolescent Sampling

The sampling method used in CHIS 2005 to select an adolescent did not change from previous cycles of CHIS. That is, an adolescent was sampled from among all eligible adolescents in a household associated with the sampled adult with equal probability. Adolescents were enumerated and sampled at the end of the screener interview if the child-first procedure was used or in section G of the adult extended interview. Since adolescents could be sampled and interviewed before the adult interview, there were some households with a completed adolescent interview where adult and/or child interviews were not completed. *CHIS 2005 Methodology Series: Report 5 - Weighting and Estimation* describes how the probabilities of selection are computed for the sampled adults, children, and adolescents.

5. ACHIEVED SAMPLE SIZES

This chapter summarizes the number of completed interviews in CHIS 2005 for the RDD strata and supplemental samples and the relationship between the targeted and the achieved numbers. As mentioned in the previous chapters, the targeted goals for CHIS 2005 were stated in terms of the total number of completed adult and child interviews obtained at the end of the data collection period. The actual number of completed interviews is a function of the number of telephone numbers sampled, the within-household person sampling, and different reasons for attrition. These reasons are discussed in more detail in Chapter 3. Detailed information about the response rates is presented in *CHIS 2005 Methodology Series: Report 4 – Response Rates*.

Table 5-1 gives the number of completed adult interviews for the RDD and geographic samples by two methods of classifying the geographic area in which the sampled adult resides. The first column of completed interviews in the table uses the data on the county that was available at the time of sampling (and during the data collection period). As noted in Chapter 3 on sampling households, each telephone number is assigned to exactly one stratum for sampling purposes, but the number may actually be for a household in a different county. The third column in the table uses the self-reported residence county of the adult respondent. This classification is based on the geocoded location of the adult's residence derived from data collected on the county, ZIP Code, address, and street intersection in the adult interview. This classification is the most appropriate for analysis of CHIS 2005 data. *CHIS 2005 Methodology Series: Report 3 – Data Processing Procedures* describes how the self-reported data were processed and how reporting discrepancies were resolved.

Table 5-1 also shows the number of completed interviews as percentages of the targeted number of adult interviews (revised targets) for the base RDD and geographic samples set at the time of the design. A percentage of 100 or greater indicates the targeted number of adult interviews was reached in the stratum. Table 5-1 shows that the overall target was not met in CHIS 2005. The targets were met or surpassed in 15 strata of the 44 strata based on the sampling location information that was available at the time of data collection. For the self-reported location, 17 of the 44 strata met or surpassed the target number of completes. The discrepancies between the two location classifications are largely a function of how well the sampling classification matched with the self-reported classification.

Table 5-1. Number of completed adult interviews for the base RDD and geographic samples by sampling and self-reported stratum*

Stratum	Sampling location		Self-reported location	
	Completed interviews	Percentage of target	Completed interviews	Percentage of target
State Total	41,074	97.9	41,074	97.9
Los Angeles	8,488	94.3	8,499	94.4
San Diego	2,639	100.6	2,642	100.8
Orange	2,377	101.8	2,330	99.7
Santa Clara	1,382	100.1	1,386	100.4
San Bernardino	1,289	93.1	1,294	93.4
Riverside	1,278	92.6	1,292	93.6
Alameda	1,287	99.9	1,255	97.4
Sacramento	1,143	96.9	1,148	97.4
Contra Costa	823	95.5	873	101.3
Fresno	587	90.9	584	90.4
San Francisco	758	94.8	744	93.0
Ventura	613	100.2	634	103.6
San Mateo	649	106.9	681	112.2
Kern	595	107.6	590	106.7
San Joaquin	450	92.2	452	92.6
Sonoma	481	98.6	489	100.2
Stanislaus	456	93.4	433	88.7
Santa Barbara	464	95.1	461	94.5
Solano	1,211	98.5	1,184	96.3
Tulare	469	96.1	481	98.6
Santa Cruz	510	104.5	496	101.6
Marin	3,103	103.4	3,089	103.0
San Luis Obispo	486	99.6	485	99.4
Placer	472	96.7	469	96.1
Merced	486	99.6	509	104.3
Butte	465	95.3	474	97.1
Shasta	498	102.0	479	98.2
Yolo	474	97.1	481	98.6
El Dorado	456	93.4	458	93.9
Imperial	423	86.7	422	86.5
Napa	475	97.3	488	100.0
Kings	466	95.5	463	94.9
Madera	477	97.7	472	95.5
Monterey	532	109.0	546	111.1

Table 5-1. Number of completed adult interviews for the base RDD and geographic samples by sampling and self-reported stratum* (Continued)

Stratum	Sampling location		Self-reported location	
	Completed interviews	Percentage of target	Completed interviews	Percentage of target
Humboldt	819	102.4	815	101.5
Nevada	403	103.3	402	102.6
Mendocino	414	106.2	408	104.1
Sutter	382	97.9	377	95.9
Yuba	378	96.9	357	91
Lake	383	98.2	377	96.2
San Benito	349	89.5	349	89.2
Colusa, Glenn, Tehama	410	105.1	428	109
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	382	97.9	394	100.8
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	392	100.5	384	98.5

*Partially completed interviews (completed through at least Section J) are counted as complete

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Similarly to Table 5-1, Table 5-2 shows the number of completed interviews as percentages of the targeted number of child interviews (revised targets) for the base RDD and geographic samples set at the time of the design. For child interviews, the overall target of children was exceeded in CHIS 2005. Targets were fully met or exceeded in 35 strata based on the sampling location and for self-reported location the target was fully met or exceeded in 33 strata.

Table 5-2. Number of completed child interviews for the base RDD and geographic samples by sampling and self-reported stratum

Stratum	Sampling location		Self-reported location	
	Completed interviews	Percentage of target	Completed interviews	Percentage of target
State Total	9,605	106.1	9,605	106.1
Los Angeles	2,002	109.2	2,009	109.6
San Diego	601	109.7	603	110.0
Orange	594	114.9	581	112.4
Santa Clara	388	127.2	387	126.9
San Bernardino	373	109.4	371	108.8
Riverside	329	100.0	331	100.6
Alameda	285	103.3	274	99.3
Sacramento	260	109.2	264	110.9
Contra Costa	186	102.8	199	109.9
Fresno	155	79.1	158	80.6
San Francisco	93	86.9	91	85.0
Ventura	161	120.1	165	123.1
San Mateo	154	131.6	157	134.2
Kern	175	129.6	176	130.4
San Joaquin	123	103.4	123	103.4
Sonoma	76	74.5	76	74.5
Stanislaus	119	100.8	112	94.9
Santa Barbara	119	103.5	117	101.7
Solano	291	104.7	286	102.9
Tulare	140	103.7	141	104.4
Santa Cruz	115	125.0	109	118.5
Marin	564	93.1	564	93.1
San Luis Obispo	93	104.5	95	106.7
Placer	107	102.9	109	104.8
Merced	155	104.7	159	107.4
Butte	70	74.5	71	75.5
Shasta	102	117.2	96	110.3
Yolo	122	114.0	126	117.8

Table 5-2. Number of completed child interviews for the base RDD and geographic samples by sampling and self-reported stratum (Continued)

Stratum	Sampling location		Self-reported location	
	Completed interviews	Percentage of target	Completed interviews	Percentage of target
El Dorado	107	107.0	103	103.0
Imperial	119	92.2	119	92.2
Napa	98	105.4	103	110.8
Kings	176	107.3	175	106.7
Madera	113	102.7	111	100.9
Monterey	145	115.1	151	119.8
Humboldt	148	98.7	148	98.7
Nevada	67	136.7	62	126.5
Mendocino	78	113.0	77	111.6
Sutter	90	92.8	88	90.7
Yuba	117	114.7	115	112.7
Lake	58	69.9	58	69.9
San Benito	111	101.8	112	102.8
Colusa, Glenn, Tehama	98	105.4	103	110.8
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	73	102.8	75	105.6
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	55	98.2	55	98.2

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Table 5-3 shows the number of completed adolescent interviews for the RDD and geographic samples. Because there were not predetermined targets by stratum for adolescents, the column for the percentages of the targeted number of interviews are not included in the table. However, we expected between 3,000 and 4,000 completed adolescent interviews and 3,739 were completed.

Table 5-3. Number of completed adolescent interviews for the base RDD and geographic samples by sampling and self-reported stratum

Stratum	Sampling location	Self-reported location
	Completed interviews	Completed interviews
State Total	3,739	3,739
Los Angeles	769	769
San Diego	222	223
Orange	206	201
Santa Clara	119	119
San Bernardino	146	147
Riverside	130	131
Alameda	100	99
Sacramento	106	108
TContra Costa	74	76
Fresno	69	69
San Francisco	32	31
Ventura	60	62
San Mateo	55	57
Kern	66	65
San Joaquin	45	43
Sonoma	31	31
Stanislaus	45	39
Santa Barbara	44	43
Solano	111	111
Tulare	49	50
Santa Cruz	53	52
Marin	246	246
San Luis Obispo	44	45
Placer	43	44
Merced	50	56
Butte	37	39
Shasta	43	38
Yolo	58	56
El Dorado	37	37
Imperial	59	59
Napa	31	32
Kings	49	50
Madera	46	45
Monterey	59	62

Table 5-3. Number of completed adolescent interviews for the base RDD and geographic samples by sampling and self-reported stratum (Continued)

Stratum	Sampling location	Self-reported location
	Completed interviews	Completed interviews
Nevada	34	34
Mendocino	31	30
Sutter	43	43
Yuba	45	41
Lake	32	31
San Benito	54	54
Colusa, Glenn, Tehama	38	43
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	33	33
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	26	26

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Table 5-4 shows the number of completed adult interviews for the Korean and Vietnamese surname list samples. The supplemental sample targets were revised during the data collection period as experience was gained on the actual RDD sample yield. The target was exceeded for the number of completed Korean adult interviews but not for the Vietnamese interviews. The yield both from the base RDD and from the Vietnamese supplemental sample was well below what was expected.

Table 5-4. Number of completed adult, child, and adolescent interviews by surname list sample

Sample	Number of completed interviews	
	Korean	Vietnamese
Base RDD	433	336
Korean list	83	1
Vietnamese list	0	72
Korean-Vietnamese list	82	31
Total	598	440
Target	500	500
Percentage of Target	119.6	88.0

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Table 5-5 shows the number of completed adult interviews for the child supplemental samples. Although the state supplement achieved only 91 percent of the target, the overall number of child interviews statewide exceeded the target as described earlier. In San Diego County, the target was

99 percent achieved, and the number of child interviews overall exceeded the target in the county. The proportion of households reporting children in the screening interview was lower for both the statewide and San Diego supplements than was anticipated. This shortfall was more than offset by the success of the child-first approach in increasing the number of completed child interviews in all samples.

Table 5-5. Number of completed adult and child interviews by type of child supplemental sample

Child supplemental samples	Number of completed interviews	
	Adult	Children
State		
Total	525	511
Target	562	562
Percent of target	93.4	90.9
San Diego		
Target	1,166	1,166
Total (sampling stratum)	1,143	1,160
Total (self-reported stratum)	1,140	1,158
Percentage of target (self-reported)	97.8	99.3

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Table A-3 through A-5 in the appendix shows the number of completed interviews by self-reported stratum for the adult, child and adolescent samples by the different sample types.

REFERENCES

- Anderson, J.E., Nelson, D.E., and Wilson, R.W. (1998). Telephone coverage and measurement of health risk indicators: Data from the National Health Interview Survey. *American Journal of Public Health*, 88, 1392-1395.
- Blumberg, S. J., Luke, J. V., and Cynamon, M. L. (2006). Telephone coverage and health survey estimates: evaluating the need for concern about wireless substitution. *American Journal of Public Health*: Vol. 96, No. 5, 926-931.
- Brick, J.M., Montaquila, J., Hagedorn, M.C., Brock Roth, S., and Chapman, C. (2005). Implications for RDD design from an incentive experiment. *Journal of Official Statistics*, 21(4), 571-589.
- Brick, J.M., and Waksberg, J. (1991). Avoiding sequential sampling with RDD. *Survey Methodology*, 17(1), 27-41.
- Brick, J.M., Waksberg, J., Kulp, D., and Starer, A. (1995). Bias in list-assisted telephone surveys. *Public Opinion Quarterly*, 59(2), 218-235.
- Elliott, M R., Little R. J.A., and Lewitzky S. (2000). Subsampling callbacks to improve survey efficiency. *Journal of the American Statistical Association*, 95:451, 730-738.
- Ford, E.S. (1998). Characteristics of survey participants with and without a telephone: Findings from the Third National Health and Nutrition Examination Survey. *Journal of Clinical Epidemiology*, 51, 55-60.
- Giesbrecht, L.H., Kulp, D.W., and Starer, A.W. (1996). Estimating coverage bias in RDD samples with current population survey data. *Proceedings of the Survey Research Methods Section*, 503-508. American Statistical Association.
- Hansen, M.H., and Hurwitz, W.N. (1946). The problem of nonresponse in sample surveys. *Journal of the American Statistical Association*, 41, 517-529.
- Kalton, G., and Anderson, D.W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society, A*, 149, 65-82.
- Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, 44, 380-87.
- Kish, L. (1992). Weighting for unequal P_i . *Journal of Official Statistics*, 8, 183-200.
- Lee, S., Kurata, J., Nguyen, H.A., and Jawad, M. (2006). *Shifting Focus from Nonresponse to Coverage Error: Linguistic Minorities in Surveys*. Presented at the annual meeting of the American Association for Public Opinion Research, Montréal, Canada.

Rizzo, L., Brick, J.M., and Park, I. (2004). A minimally intrusive method for sampling persons in Random Digit Dialing Surveys. *Public Opinion Quarterly*, 68, 267-274.

Sudman, S., Sirken, M.G., and Cowan, C.D. (1988). Sampling rare and elusive populations. *Science*, 240, 991-9.

Tucker, C., Brick, J.M., Meekins, B., and Morganstein, D. (2004). Household telephone service and usage patterns in the U.S. in 2004. *Proceedings of the Survey Methods Section of the American Statistical Association*: CD-ROM.

Tucker, C., Lepkowski, J., and Piekarski, L. (2002). The current efficiency of list-assisted telephone sampling designs. *Public Opinion Quarterly*, 66, 321-338.

APPENDIX

Table A-1. Stratum definitions for CHIS 2001, 2003 and 2005

County	2005 Stratum	2001 and 2003 Stratum
Los Angeles	1	1
San Diego	2	2
Orange	3	3
Santa Clara	4	4
San Bernardino	5	5
Riverside	6	6
Alameda	7	7
Sacramento	8	8
Contra Costa	9	9
Fresno	10	10
San Francisco	11	11
Ventura	12	12
San Mateo	13	13
Kern	14	14
San Joaquin	15	15
Sonoma	16	16
Stanislaus	17	17
Santa Barbara	18	18
Solano	19	19
Tulare	20	20
Santa Cruz	21	21
Marin	22	22
San Luis Obispo	23	23
Placer	24	24
Merced	25	25
Butte	26	26
Shasta	27	27
Yolo	28	28
El Dorado	29	29
Imperial	30	30
Napa	31	31
Kings	32	32
Madera	33	33
Monterey	34	34
San Benito	41	34
Lake	40	37
Mendocino	37	37
Sutter	38	39
Yuba	39	39

Table A-1. Stratum definitions for CHIS 2001, 2003 and 2005 (continued)

County	2005 Stratum	2001 and 2003 Stratum
Colusa Glen Tehama	42	38
Humboldt Del Norte,	35	35
Lassen Modoc Siskiyou Trinity	43	36
Plumas Sierra Nevada	36	40
Alpine Amador Calaveras Inyo Mariposa Mono Tuolumne	44	41

Table A-2. Number of telephone numbers drawn by sample type and sampling stratum

Stratum	Base RDD	Supplemental sample						Total
		Geographic	List*			Child		
			Korean	Vietnamese	Korean/ Vietnamese	State*	San Diego	
State	432,446	35,175	1,832	1,438	1,608	23,787	0	554,572
1 Los Angeles	113,627	0	694	307	560	6,669	0	121,857
2 San Diego	30,886	0	65	90	66	1,890	0	91,283
3 Orange	35,931	0	184	326	232	2,136	0	38,809
4 Santa Clara	19,728	0	189	244	178	1,237	0	21,576
5 San Bernardino	13,207	0	41	32	40	989	0	14,309
6 Riverside	12,656	0	29	35	34	977	0	13,731
7 Alameda	15,512	0	132	85	99	1,239	0	17,067
8 Sacramento	11,743	0	48	68	56	914	0	12,829
9 Contra Costa	8,647	0	42	20	31	762	0	9,502
10 Fresno	5,937	0	18	13	19	487	0	6,474
11 San Francisco	12,257	0	179	85	114	891	0	13,526
12 Ventura	7,142	0	18	9	16	454	0	7,639
13 San Mateo	9,561	0	60	20	42	623	0	10,306
14 Kern	6,496	0	7	4	7	386	0	6,900
15 San Joaquin	3,898	0	15	21	15	297	0	4,246
16 Sonoma	3,883	0	10	8	7	322	0	4,230
17 Stanislaus	3,518	0	5	5	5	232	0	3,765
18 Santa Barbara	6,058	0	8	6	5	276	0	6,353
19 Solano	10,965	700	8	5	10	214	0	11,902
20 Tulare	4,743	0	3	2	1	203	0	4,952
21 Santa Cruz	4,671	0	5	4	3	191	0	4,874
22 Marin	5,164	32,575	8	5	9	264	0	38,025
23 San Luis Obispo	4,321	0	5	4	6	171	0	4,507
24 Placer	4,599	0	7	3	10	218	0	4,837
25 Merced	3,800	0	3	1	2	85	0	3,891
26 Butte	2,899	0	3	3	1	115	0	3,021
27 Shasta	3,277	0	2	3	2	110	0	3,394
28 Yolo	3,967	0	15	8	10	104	0	4,104
29 El Dorado	4,122	0	3	2	2	99	0	4,228
30 Imperial	4,630	0	0	0	2	60	0	4,692
31 Napa	5,210	0	1	2	3	83	0	5,299
32 Kings	3,700	0	1	2	0	47	0	3,750
33 Madera	3,900	0	0	1	0	60	0	3,961
34 Monterey	7,418	0	11	6	7	263	0	7,705
35 Humboldt	4,297	1,900	2	1	2	86	0	6,288
36 Nevada	2,868	0	2	1	2	69	0	2,942
37 Mendocino	3,200	0	2	1	2	61	0	3,266
38 Sutter	3,442	0	1	1	1	35	0	3,480
39 Yuba	3,054	0	1	0	1	36	0	3,092
40 Lake	3,063	0	0	0	0	44	0	3,107

Table A-2. Number of telephone numbers drawn by sample type and sampling stratum (continued)

	Stratum	Base RDD	Geographic	Supplemental sample					Total
				List*			Child		
				Korean	Vietnamese	Korean/ Vietnamese	State*	San Diego	
41	San Benito	3,798	0	0	1	1	25	0	3,825
42	Colusa, Glenn, Tehama	2,921	0	1	1	0	55	0	2,978
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	3,984	0	1	0	2	136	0	4,123
44	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	3,746	0	3	3	3	172	0	3,927

* Not drawn by sampling stratum.

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Table A-3. Number of adult completed interviews by self-reported stratum

Stratum	Base RDD	Supplemental sample						Total
		Geographic	List*			Child		
			Korean	Vietnamese	Korean/ Vietnamese	State*	San Diego	
State	38,089	2,985	84	79	115	525	1,143	43,020
1 Los Angeles	8,499	0	46	10	49	117	1	8,722
2 San Diego	2,642	0	3	7	3	33	1,140	3,828
3 Orange	2,330	0	16	27	27	47	0	2,447
4 Santa Clara	1,386	0	6	20	7	26	0	1,445
5 San Bernardino	1,294	0	2	1	1	33	0	1,331
6 Riverside	1,292	0	2	2	5	35	1	1,337
7 Alameda	1,253	2	1	2	1	26	0	1,285
8 Sacramento	1,148	0	1	4	3	15	0	1,171
9 Contra Costa	872	1	1	1	2	17	0	894
10 Fresno	584	0	0	0	0	11	0	595
11 San Francisco	737	7	1	2	5	11	0	763
12 Ventura	634	0	0	1	3	15	1	654
13 San Mateo	681	0	1	0	2	11	0	695
14 Kern	590	0	0	0	1	10	0	601
15 San Joaquin	452	0	0	1	0	18	0	471
16 Sonoma	484	5	0	0	2	5	0	496
17 Stanislaus	433	0	0	0	1	10	0	444
18 Santa Barbara	461	0	0	0	0	8	0	469
19 Solano	1,116	68	0	0	0	5	0	1,189
20 Tulare	481	0	0	0	0	4	0	485
21 Santa Cruz	496	0	1	0	0	5	0	502
22 Marin	458	2,631	1	0	1	6	0	3,097
23 San Luis Obispo	485	0	0	0	0	5	0	490
24 Placer	468	1	0	0	1	1	0	471
25 Merced	509	0	0	0	0	4	0	513
26 Butte	474	0	0	0	0	2	0	476
27 Shasta	478	1	0	0	0	3	0	482
28 Yolo	480	1	0	0	1	4	0	486
29 El Dorado	458	0	0	0	0	3	0	461
30 Imperial	422	0	0	0	0	3	0	425
31 Napa	485	3	0	1	0	1	0	490
32 Kings	463	0	0	0	0	3	0	466
33 Madera	472	0	0	0	0	1	0	473
34 Monterey	546	0	1	0	0	7	0	554
35 Humboldt	557	258	0	0	0	3	0	818
36 Nevada	400	2	0	0	0	0	0	402
37 Mendocino	407	1	0	0	0	3	0	411
38 Sutter	377	0	1	0	0	2	0	380

Table A-3. Number of adult completed interviews by self-reported stratum (continued)

Stratum	Base RDD	Supplemental sample						Total
		Geographic	List *			Child		
			Korean	Vietnamese	Korean/ Vietnamese	State *	San Diego	
39 Yuba	357	0	0	0	0	0	0	357
40 Lake	377	0	0	0	0	1	0	378
41 San Benito	349	0	0	0	0	2	0	351
42 Colusa, Glenn, Tehama	428	0	0	0	0	2	0	430
43 Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	390	4	0	0	0	2	0	396
44 Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	384	0	0	0	0	5	0	389

* Not drawn by sampling stratum.

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Table A-4. Number of child completed interviews by self-reported stratum

Stratum	Base RDD	Supplemental sample						Total
		Geographic	List*			Child		
			Korean	Vietnamese	Korean/ Vietnamese	State*	San Diego	
State	9,066	539	22	22	38	511	1,160	11,358
1 Los Angeles	2,009	0	7	6	17	110	1	2,150
2 San Diego	603	0	1	2	0	33	1,158	1,797
3 Orange	581	0	7	3	8	49	0	648
4 Santa Clara	387	0	2	7	0	27	0	423
5 San Bernardino	371	0	0	0	1	28	0	400
6 Riverside	331	0	1	1	3	33	1	370
7 Alameda	274	0	0	0	0	27	0	301
8 Sacramento	264	0	1	1	0	17	0	283
9 Contra Costa	199	0	0	1	1	16	0	217
10 Fresno	158	0	0	0	0	14	0	172
11 San Francisco	91	0	0	1	2	8	0	102
12 Ventura	165	0	0	0	2	12	0	179
13 San Mateo	157	0	0	0	1	12	0	170
14 Kern	176	0	0	0	1	6	0	183
15 San Joaquin	123	0	0	0	0	18	0	141
16 Sonoma	76	0	0	0	1	6	0	83
17 Stanislaus	112	0	0	0	1	8	0	121
18 Santa Barbara	117	0	0	0	0	7	0	124
19 Solano	264	22	0	0	0	5	0	291
20 Tulare	141	0	0	0	0	5	0	146
21 Santa Cruz	109	0	0	0	0	6	0	115
22 Marin	89	475	1	0	0	8	0	573
23 San Luis Obispo	95	0	0	0	0	5	0	100
24 Placer	109	0	0	0	0	3	0	112
25 Merced	159	0	0	0	0	4	0	163
26 Butte	71	0	0	0	0	3	0	74
27 Shasta	96	0	0	0	0	3	0	99
28 Yolo	126	0	0	0	0	3	0	129
29 El Dorado	103	0	0	0	0	3	0	106
30 Imperial	119	0	0	0	0	3	0	122
31 Napa	103	0	0	0	0	3	0	106
32 Kings	175	0	0	0	0	2	0	177
33 Madera	111	0	0	0	0	1	0	112
34 Monterey	151	0	1	0	0	6	0	158
35 Humboldt	106	42	0	0	0	3	0	151
36 Nevada	62	0	0	0	0	0	0	62
37 Mendocino	77	0	0	0	0	1	0	78
38 Sutter	88	0	1	0	0	1	0	90

Table A-4. Number of child completed interviews by self-reported stratum (continued)

Stratum	Base RDD	Supplemental sample						Total
		Geographic	List*			Child		
			Korean	Vietnamese	Korean/ Vietnamese	State*	San Diego	
39 Yuba	115	0	0	0	0	0	0	115
40 Lake	58	0	0	0	0	1	0	59
41 San Benito	112	0	0	0	0	2	0	114
42 Colusa, Glenn, Tehama	103	0	0	0	0	2	0	105
43 Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	75	0	0	0	0	3	0	78
44 Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	55	0	0	0	0	4	0	59

* Not drawn by sampling stratum.

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.

Table A-5. Number of adolescent completed interviews by self-reported stratum

Stratum	Base RDD	Supplemental sample						Total
		Geographic	List*			Child		
			Korean	Vietnamese	Korean/ Vietnamese	State*	San Diego	
State	3,497	242	6	6	8	84	186	4,029
1 Los Angeles	769	0	3	2	1	17	0	792
2 San Diego	223	0	0	0	0	3	186	412
3 Orange	201	0	0	4	4	8	0	217
4 Santa Clara	119	0	0	0	0	6	0	125
5 San Bernardino	147	0	0	0	0	4	0	151
6 Riverside	131	0	0	0	1	6	0	138
7 Alameda	99	0	1	0	0	5	0	105
8 Sacramento	108	0	0	0	0	2	0	110
9 Contra Costa	76	0	0	0	0	2	0	78
10 Fresno	69	0	0	0	0	4	0	73
11 San Francisco	31	0	1	0	0	0	0	32
12 Ventura	62	0	0	0	1	3	0	66
13 San Mateo	57	0	1	0	0	1	0	59
14 Kern	65	0	0	0	0	1	0	66
15 San Joaquin	43	0	0	0	0	3	0	46
16 Sonoma	31	0	0	0	1	2	0	34
17 Stanislaus	39	0	0	0	0	2	0	41
18 Santa Barbara	43	0	0	0	0	3	0	46
19 Solano	103	8	0	0	0	3	0	114
20 Tulare	50	0	0	0	0	1	0	51
21 Santa Cruz	52	0	0	0	0	0	0	52
22 Marin	36	210	0	0	0	0	0	246
23 San Luis Obispo	45	0	0	0	0	0	0	45
24 Placer	44	0	0	0	0	0	0	44
25 Merced	56	0	0	0	0	0	0	56
26 Butte	39	0	0	0	0	0	0	39
27 Shasta	38	0	0	0	0	2	0	40
28 Yolo	56	0	0	0	0	1	0	57
29 El Dorado	37	0	0	0	0	2	0	39
30 Imperial	59	0	0	0	0	0	0	59
31 Napa	32	0	0	0	0	1	0	33
32 Kings	50	0	0	0	0	1	0	51
33 Madera	45	0	0	0	0	0	0	45
34 Monterey	62	0	0	0	0	0	0	62
35 Humboldt	45	24	0	0	0	0	0	69
36 Nevada	34	0	0	0	0	0	0	34
37 Mendocino	30	0	0	0	0	0	0	30

Table A-5. Number of adolescent completed interviews by self-reported strata (continued)

Stratum	Base RDD	Supplemental sample						Total
		Geographic	List*			Child		
			Korean	Vietnamese	Korean/ Vietnamese	State*	San Diego	
38 Sutter	43	0	0	0	0	0	0	43
39 Yuba	41	0	0	0	0	0	0	41
40 Lake	31	0	0	0	0	0	0	31
41 San Benito	54	0	0	0	0	0	0	54
42 Colusa, Glenn, Tehama	43	0	0	0	0	0	0	43
43 Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	33	0	0	0	0	1	0	34
44 Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	26	0	0	0	0	0	0	26

* Not drawn by sampling stratum.

Source: UCLA Center for Health Policy Research, 2005 California Health Interview Survey.