November, 2011

CHIS 2009 Methodology Report Series

Report 1

Sample Design

CHIS 2009 METHODOLOGY SERIES

REPORT 1

SAMPLE DESIGN

NOVEMBER 2011

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 2009, 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 2009 Methodology Series: Report 1 - Sample Design*. Los Angeles, CA: UCLA Center for Health Policy Research, 2011.

Copyright © 2011 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 Public Health, and the Department of Health Care Services. Funding for CHIS 2009 came from multiple sources: the California Department of Public Health, the Department of Health Care Services, the California Endowment, the National Cancer Institute, NIH Office of Behavioral and Social Sciences Research, Agency for Healthcare Research and Quality, the California Wellness Foundation, First 5 California, the California Department of Mental Health, the California Office of the Patient Advocate, Kaiser Permanente, Blue Shield of California Foundation, the California HealthCare Foundation, the San Diego County Human and Health Services Agency, Marin County Department of Health and Human Services, and the Humboldt County Department of Health and Human Services.

PREFACE

Sample Design is the first in a series of methodological reports describing the 2009 California Health Interview Survey (CHIS 2009). The other reports are listed below.

CHIS is a collaborative project of the University of California, Los Angeles (UCLA) Center for Health Policy Research, the California Department of Public Health, and the Department of Health Care Services. Westat was responsible for the data collection and the preparation of five methodological reports for the 2009 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 2009 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 2009. An appropriate sample design is a feature of a successful survey, and CHIS 2009 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.

This report provides analysts information about the sampling methods used for CHIS 2009, 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

TABLE OF CONTENTS (CONTINUED)

List of Tables

TABLE OF CONTENTS (CONTINUED)

Appendix

1. CHIS 2009 SAMPLE DESIGN AND METHODOLOGY SUMMARY

1.1 Overview

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

The sample is designed to meet and optimize two objectives:

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

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

This series of reports describes the methods used in collecting data for CHIS 2009, the fifth CHIS data collection cycle, which was conducted between September 2009 and April 2010. The previous CHIS cycles (2001, 2003, 2005, and 2007) are described in similar series, available at http://www.chis.ucla.edu/methods.html.

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

1.2 Sample Design Objectives

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

Table 1-1 shows the 44 sampling strata, which include 41 independent county strata. A sufficient number of adult interviews were allocated to each stratum to support the first sample design objective—to provide health estimates for adults at the local level. The geographic stratification of the state was the same as that used since CHIS 2005. In the first two CHIS cycles there were 41 total sampling strata, including 33 individual counties. The CHIS 2009 samples in Humboldt, Marin, and San Diego Counties were enhanced with additional funding.

The main landline RDD CHIS sample size is sufficient to accomplish the second objective. To increase the precision of estimates for Koreans and Vietnamese, areas with relatively high concentrations of these groups were sampled at higher rates. These geographically targeted oversamples were supplemented by telephone numbers associated with group-specific surnames drawn from listed telephone directories to further increase the sample size for Koreans and Vietnamese. CHIS 2009 included additional Korean and Vietnamese oversamples conducted on behalf of the National Cancer Institute.

To help compensate for the increasing number of households without landline telephone service, a separate RDD sample was drawn of telephone numbers assigned to cellular service. In CHIS 2009, the goal was to complete approximately 2,500 interviews statewide with adults from the cell-phone sample . The CHIS 2009 cell-phone sampled from the CHIS 2007 cell-phone sample in two significant ways. First, all cell-phone sample cases were eligible for the extended interview regardless of the presence of a landline phone. The landline and cell samples, therefore, overlap and contrasts to CHIS 2007 when cell-phone cases with a landline telephone were screened out to limit the cell-phone sample to "cell-phone only" cases. This change was made due to the large and potentially unique characteristics of

telephone users who possess both a landline and cell-phone, but rely principally on their cell-phone for communication and would otherwise be excluded from the sample. The second change to the cell-phone sample was the inclusion of child and adolescent extended interviews. About 200 teen interviews and nearly 500 child interviews were completed from the cell-phone sample in CHIS 2009. Because data are not available for numbers assigned to cellular service to support the same level of geographic stratification as the landline sample, the cell RDD sample was stratified by area code. If the sampled number was shared by two or more adult members of a cell-only household, one household member was selected for the adult interview. Otherwise, the adult owner of the sampled number was selected.

Table 1-1. California county and county group strata used in the CHIS 2009 sample design

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

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

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 2009 data collection under contract with the UCLA Center for Health Policy Research. For the landline RDD sample, Westat staff interviewed one randomly selected adult in each sampled household, and sampled one adolescent and one child if present in the household and the sampled adult was the parent or legal guardian. Up to three interviews could have been completed in each household. In households with children where the sampled adult was not the screener respondent, children and adolescents could be sampled as part of the screening interview, and the extended child (and adolescent) interviews could be completed before the adult interview. This "child-first" procedure was new for CHIS 2005 and has been continued in subsequent CHIS cycles; this procedure substantially increases 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 2009 by the type of sample (landline RDD, surname list, and cell RDD).

Table 1-2. Number of completed CHIS 2009 interviews by type of sample and instrument

Type of sample	Adult	Child	Adolescent
Total all samples	47,614	8,945	3,379
Landline RDD	42,682	7,918	3,002
Surname list	1,885	545	178
Cell RDD	3,047	482	199

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Interviews in all languages were administered using Westat's computer-assisted telephone interviewing (CATI) system. The average adult interview took about 40 minutes to complete. The average child and adolescent interviews took about 16 minutes and 18 minutes, respectively. For "child-first" interviews, additional household information asked as part of the child interview averaged about 9 minutes. Interviews in non-English languages generally took longer to complete. More than 12 percent of the adult interviews were completed in a language other than English, as were almost 24 percent of all child (parent proxy) interviews and 9 percent of all adolescent interviews.

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

1.4 Response Rates

The overall response rate for CHIS 2009 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 landline sampled telephone numbers for which an address could be obtained from reverse directory services. An advance letter was mailed for approximately 58 percent of the landline RDD sample telephone numbers, and 82 percent of list sample numbers. Addresses were not available for the cell sample. As in CHIS 2005 and 2007, a $2 bill was included with the advance letter to promote cooperation.

The CHIS 2009 screener completion rate for the landline and samples was 36.1 percent, and was higher for households that were sent the advance letter. For the cell phone sample, the screener completion rate was 19.3 percent in all households. The extended interview completion rate for the landline sample varied across the adult (49.0 percent), child (72.9 percent) and adolescent (42.8 percent) interviews. The adolescent rate includes getting permission from a parent or guardian. The adult interview completion rate for the cell sample was 56.2 percent. Multiplying the screener and extended rates gives an overall response rate for each type of interview. The percentage of households completing one or more of the extended interviews (adult, child, and/or adolescent) is a useful summary of the overall performance of the landline sample. For CHIS 2009, the landline sample household response rate was 19.7 percent (the product of the screener response rate and the completion rate at the household level of 54.7 percent). All of the household and person level response rates vary by sampling stratum. For more information about the CHIS 2009 response rates, please see *CHIS 2009 Methodology Series: Report 4 – Response Rates*.

Table 1-3. CHIS 2009 survey topic areas by instrument

Health status	**Adult**	**Teen**	**Child**
General health status, height and weight	✓	✓	✓
Days missed from school due to health problems	✓	✓	✓
Health conditions	**Adult**	**Teen**	**Child**
Asthma	✓	✓	✓
Diabetes, gestational diabetes, pre-diabetes/borderline	✓		
Heart disease, high blood pressure	✓		
Physical disability	✓		
Developmental assessment and developmental conditions			✓
Mental health	**Adult**	**Teen**	**Child**
Mental health status	✓	✓	✓
Perceived need, access and utilization of mental health services	✓	✓	✓
Suicide ideation and attempts	✓		
Health behaviors	**Adult**	**Teen**	**Child**
Dietary intake, fast food, high sugar diet	✓	✓	✓
Physical activity and exercise	✓	✓	✓
Walking for transportation and leisure	✓		
Sedentary time		✓	✓
Flu Shot	✓		✓
Alcohol and tobacco use	✓	✓	
Illegal drug use		✓	
Sexual behavior	✓	✓	
HIV/STI testing		✓	
Sun exposure	✓	✓	
Women's health	**Adult**	**Teen**	**Child**
Mammography screening, hormone replacement therapy	✓		
Age at menarche, live births, menopause, birth control medications	✓		
Pregnancy status	✓	✓	
Cancer history and prevention	**Adult**	**Teen**	**Child**
Family history	✓		
Colorectal cancer screening, prostate specific antigen (PSA) test	✓		
Dental health	**Adult**	**Teen**	**Child**
Last dental visit, main reason haven't visited dentist		✓	✓

Table 1-3. CHIS 2009 survey topic areas by instrument (Continued)

Food environment	**Adult**	**Teen**	**Child**
Availability of food in household over past 12 months	✓		
Brought lunch to school from home		✓	
Doctor discussed nutrition/physical activity		✓	✓
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 and medical care)	✓	✓	✓
Medical home	✓	✓	✓
Communication problems with doctor	✓		
Long-term care	✓		
Health insurance	**Adult**	**Teen**	**Child**
Current insurance coverage, spouse's coverage, who pays for coverage	✓	✓	✓
Health plan enrollment, characteristics and plan assessment	✓	✓	✓
Employer offers coverage, respondent/spouse eligibility	✓		
Coverage over past 12 months, reason for lack of insurance	✓	✓	✓
Medical debt, high deductible health plans	✓	✓	✓
Partial scope Medi-Cal, Medi-Cal deficit reduction act requirements	✓		
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	✓		
Medi-Cal and healthy families eligibility	✓	✓	✓
Reason for Medi-Cal non-participation among potential beneficiaries	✓	✓	✓
Neighborhood and housing	**Adult**	**Teen**	**Child**
Neighborhood safety, use of parks		✓	✓
Homeownership, length of time at current residence	✓		
Civic engagement		✓	✓
Social cohesion			✓
Emergency Preparedness	**Adult**	**Teen**	**Child**
Medication supply and basic preparedness	✓		
Interpersonal Violence	**Adult**	**Teen**	**Child**
Intrapersonal violence	✓		

Table 1-3. CHIS 2009 survey topic areas by instrument (Continued)

Parental involvement/adult supervision	**Adult**	**Teen**	**Child**
Adult presence after school/knowledge of teen's activities, role models		✓	
Parental concerns/involvement			✓
Child care and school attendance	**Adult**	**Teen**	**Child**
Current child care arrangements			✓
Paid child care	✓		
First 5 California: Parent kit, educational TV programming			✓
Preschool/school attendance, name of school		✓	✓
Employment	**Adult**	**Teen**	**Child**
Employment status, spouse's employment status	✓		
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**
Race and ethnicity, age, gender, height, weight, education	✓	✓	✓
Veteran status	✓		
Marital status, registered domestic partner status	✓		
Sexual orientation	✓	✓	
Language spoken with peers, language of TV, radio, newspaper used	✓		
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, 2009 California Health Interview Survey.

Historically, the CHIS response rates are comparable to response rates of other scientific telephone surveys in California, such as the California Behavioral Risk Factor Surveillance System (BRFSS) Survey. It has become increasingly difficult, however, to compare the CHIS and BRFSS response rates due to changes in the BRFSS response rate calculation methods. 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. The 2009 BRFSS, for example, shows the refusal rate for the California (32.2%) is the highest in the nation and more than twice the national median (15.7%).[1] Survey response rates tend to

[1] As reported in the Behavioral Risk Factor Surveillance System 2009 Summary Data Quality Report (Version #1 – Revised: 04/27/2010, available online at ftp://ftp.cdc.gov/pub/Data/Brfss/2009_Summary_Data_Quality_Report.pdf

be lower in California than nationally, and over the past decade response rates have been declining both nationally and in California. Further information about CHIS data quality and nonresponse bias is available at http://www.chis.ucla.edu/dataquality.html.

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

Proxy interviews were allowed for frail and ill persons over the age of 65 who were unable to complete the extended adult interview in order to avoid biases for health estimates of elderly persons that might otherwise result. Eligible selected persons were recontacted and offered a proxy option. For 283 elderly adults, a proxy interview was completed by either a spouse/partner or adult child. A reduced questionnaire, with questions identified as appropriate for a proxy respondent, was administered. (Note: questions not administered in proxy interviews are given a value of "-2" in the data files.)

1.5 Weighting the Sample

To produce population estimates from the CHIS data, weights are applied to the sample data to compensate for the probability of selection and a variety of other factors, some directly resulting from the design and administration of the survey. The sample is weighted to represent the non-institutionalized population for each sampling stratum and statewide. The weighting procedures used for CHIS 2009 accomplish the following objectives:

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

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

Population control totals of the number of persons by age, race, and sex at the stratum level for CHIS 2009 were created primarily from the California Department of Finance's 2009 Population Estimates and 2009 Population Projections. The raking procedure used 11 raking dimensions, which are combinations of demographic variables (age, sex, race, and ethnicity), geographic variables (county, Service Planning Area in Los Angeles County, and Health Region in San Diego County), household composition (presence of children and adolescents in the household), and socio-economic variables (home ownership and education). The socio-economic variables are included to reduce biases associated with differential response rates from households with and without landline telephones. One limitation of using Department of Finance data is that it includes about 2.4 percent of the population of California who live in "group quarters" (i.e., persons living with nine or more unrelated persons). These persons were excluded from the CHIS target population and as a result, the number of persons living in group quarters was estimated and removed from the Department of Finance control totals prior to raking.

1.6 Imputation Methods

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

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

missing responses. With a hot deck, a value reported by a respondent for a particular item is assigned or donated to a "similar" person who did not respond to that item. The characteristics defining "similar" vary for different variables. To carry out hot deck imputation, the respondents to a survey item form a pool of donors, while the nonrespondents are a group of recipients. A recipient is matched to the subset pool of donors based on household and individual characteristics. A value for the recipient is then randomly imputed from one of the donors in the pool. Once a donor is used, it is removed from the pool of donors for that variable. Hot deck imputation was used to impute the same items in CHIS 2003, CHIS 2005, CHIS 2007, and CHIS 2009 (i.e., race, ethnicity, home ownership, and education).

UCLA-CHPR imputed missing values for nearly every variable in the data files other than those imputed by Westat and some sensitive variables in which nonresponse had its own meaning. Overall, item nonresponse rates in CHIS 2009 were low, with most variables missing valid responses for less than 2% of the sample. However, there were a few exceptions where item nonresponse rate was greater than 25% such as household income.

The imputation process conducted by UCLA-CHPR started with data editing, sometimes referred to as logical or relational imputation: for any missing value, a valid replacement value was sought based on known values of other variables of the same respondent or other sample(s) from the same household. For the remaining missing values, model-based 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 generalized linear model with a set of control variables (predictors). The link function of the model is corresponding to the nature of the variable being imputed, e.g. generalized linear regression for continuous variables, logistic regression for binary and multinomial variables, and negative binomial regression for counts variables. The donors and recipients are grouped based on their predicted values from the model.

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

The imputation order of the other variables followed the questionnaire. After all imputation

procedures were complete, every step in the data quality control process is 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 2009:

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

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

2. TELEPHONE SAMPLING METHODS

This chapter describes the sampling methods used in the CHIS 2009 telephone survey. CHIS 2009 consisted of a base landline random digit dialing (RDD) sample[2], three supplemental geographic samples, a Korean and Vietnamese surname list samples, and a statewide RDD cell phone sample. The base landline sample, geographic supplemental sample, and cell phone sample were drawn using RDD approaches while the list samples were drawn from separate surname lists of telephone numbers.

The first section describes the list-assisted RDD sampling methodology for the base landline sample. It also discusses some sources of undercoverage associated with landline telephone samples, such as persons who cannot be interviewed because of language limitations.

The second section describes problems associated with the increasing noncoverage of landline samples due to greater reliance on cellular telephone use and a drop in landline telephone services. The cell phone sample in CHIS 2009 addressed this problem by sampling and contacting numbers assigned to cellular service.

The third section describes the tritone and business purges of unproductive numbers implemented to reduce the number of calls to sampled but ineligible telephone numbers.

The last section reviews the supplemental samples in CHIS 2009. In order to increase the sample size for Koreans and Vietnamese, geographic areas with high concentrations of these populations were targeted in the landline sample. In addition, lists of surnames were used to supplement the landline sample. CHIS 2009 also included geographic supplemental samples for San Diego County, Marin County, and Humboldt County to increase the sample size and precision of county-level estimates.

2.1 List-Assisted Random Digit Dial Sampling of Landlines

List-assisted RDD sampling is currently the standard method for telephone surveys and has been the primary sampling method for all cycles of CHIS. This method was designed to produce an

[2] Supplemental samples selected by taking proportionally larger samples in certain geographic areas are part of the landline RDD sample.

unclustered sample that has good operational features (Tucker, Lepkowski, and Piekarski, 2002). In 100 series list-assisted sampling, the set of all telephone numbers in operating telephone prefixes is composed of 100-banks, each containing 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. Initially, this method had a small amount of noncoverage because telephone numbers in 100-banks with no listed telephone numbers (i.e., zero banks) were not sampled. Brick et al. (1995) showed that the bias from this approach was negligible for most estimates.

Changes in the structure of the U.S. telecommunications industry and an increasing number of residential exchanges have had a large impact on the 100 series list-assisted methodology. Fahimi et al. (2008) found that the exclusion of 100-banks without any listed telephone number could result in coverage losses of up to 20 percent of the households with a landline. Although there is no current information on the characteristics of the households, it is likely that these households have different characteristics from those in the traditionally sampled banks. Although the CHIS 2009 landline sample does not have a specific method to address this undercoverage directly, the weighting methods using control totals representing the entire population in California should mitigate its effects.

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 typically treated as nonresponse, but could 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 2009 and previous cycles, significant efforts have been made to limit this potential bias by interviewing in multiple languages (Lee et al., 2008). In CHIS 2009, interviews were conducted in five languages: English, Spanish, Korean, Vietnamese, and Chinese (Cantonese and Mandarin dialects). This effort eliminates a potentially large source of the bias that might result if interviewers had only been conducted in English.

2.2 Households without Landline Telephones

In most telephone surveys, households with no access to landline telephones (households with only cellular telephones and households with no telephone service of any type) are not sampled. For estimates correlated with socioeconomic measures such as health insurance coverage, food security, and

poverty, this coverage loss could introduce biases. The bias depends on the number of households with no landline telephones and the difference in characteristics of persons in households with and without a landline telephone.

Households with only cellular service account for the largest proportion of those without a landline. The numbers of households and persons in the United States who have cell phones have greatly increased in the last few years. The most recent estimate of cell-phone-only households is 26.6 percent nationally for the last 6 months of 2010 (Blumberg and Luke, 2010). Blumberg and Luke (2010) also reported that a sizeable proportion of households may be difficult to reach even though they have a landline because they rely on cell phones for most of their calls. This source of bias is likely to continue to grow along with the prevalence of cell phones.

The characteristics of persons in cell-phone-only households are different from those in households with landlines. For example, the cell-phone-only adults were much less likely to be insured than the adults in households with landlines. Demographic differences such as age and gender are also associated with cell-phone-only households, where the younger and males are more likely to live in cell-only households. Additionally, adults living in cell-only households are more likely than those in households with landlines to be renters or living with unrelated adults. Since this population is excluded from landline telephone surveys, there is increasing concern about the quality of estimates. For example, some observed decreases in certain prevalence measures among young adults are thought to be the result of undercoverage of young adults in cell-phone-only households (Delnevo et al., 2008). Such findings suggest that bias due to the failure to cover these households is possible.

CHIS 2009 included a cell phone sample component that addresses the potential biases from excluding cell phone only households. The feasibility of a cell phone sample was evaluated in 2005 with a pilot study of cell phone numbers in CHIS 2005 to collect data from adults in cell-only households (Brick, Edwards, and Lee 2007). This line of research was expanded to an operational statewide adult cell phone sample allocated by regions as an additional component to CHIS. Unlike previous cycles, the CHIS 2009 cell sample collected information from households with landlines who were contacted through the cell phone sample. The cell phone sample was also used to collect information on children and adolescents for the first time in CHIS. Additional details on the selection of this sample are presented in Section 3.2.

2.3 Increasing the Efficiency of Data Collection

When landline telephone numbers are sampled, special procedures are often implemented before data collection to reduce costs and to increase the efficiency of sampling and data collection. These techniques had been used in previous cycles of CHIS.

In CHIS 2009 we used tritone tests (the distinctive three-bell sound heard when dialing a nonworking number) and business purge methods to reduce the number of unproductive numbers (i.e., business and nonworking numbers). The procedure, called Comprehensive Screening Service (CSS), is offered by Market Systems Group (MSG), the vendor that also provided the sampling frames for CHIS. In the CSS, telephone numbers are matched to numbers in the White and Yellow Pages to identify nonresidential business numbers. A second procedure is a tritone test to identify nonworking numbers; a telephone number is classified as nonresidential if a tritone is encountered in two separate tests. All numbers, including those listed in the White Pages, were included in the tritone test. The method also identifies cell phone numbers that were ported from landline exchanges; in CHIS 2009, these numbers were included in the cell sample.

Table 2-1 shows the CSS result codes as well as the distribution of the sampled telephone numbers in CHIS 2009. Approximately 49 percent of the sampled numbers (CSS result codes LB, FM, NR NW, and some UB) were excluded from dialing. This was 2 percentage points higher than the 47 percent purged in CHIS 2007.

Table 2-1. CSS result codes and their distribution in the CHIS 2009 sample

CSS result code	Description	Number of telephones	Percentage
CP	Agent identified cell phone	208	0.02
DK	Undetermined	326,682	33.62
FM	Fax/modem	35,731	3.68
LA	Language barrier	8,994	0.93
LB	Listed business	34,475	3.55
NR	No ring-back	9,374	0.96
NW	Nonworking	352,873	36.32
PM	Privacy manager	7,772	0.80
RS	Residence	146,346	15.06
UB	Unlisted business	46,485	4.78
WR	Wireless number	2,681	0.28
Total		971,621	100.00

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

2.4 Supplemental Sampling

The first type of supplemental sample implemented in CHIS 2009 was geographic sampling to increase the sample size in specified geographic areas. CHIS 2009 included supplemental samples in San Diego County, Humboldt County, and Marin County. The selection of these samples is described in Section 3.2.

The second type of supplemental sampling in CHIS was used to improve the sample size and precision of the estimates for specific race and ethnic groups. As mentioned in Chapter 1, one of the goals of CHIS 2009 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. The expected sample yield from the landline sample was too small to support inferences for these groups at the desired level of precision. Since CHIS 2003, two sampling strategies have been used to meet a target sample yield of 500 for Korean and 500 Vietnamese adult interviews (Edwards et al., 2002). In CHIS 2009, the targets were increased to 1,000 Korean and 1,000 Vietnamese adult interviews to support testing of the discrimination module with these groups. The sampling strategies used to increase the sample sizes for these groups are disproportionate stratified sampling and multiple frame sampling. These strategies are mainly used to oversample rare or small populations (Flores Cervantes and Kalton, 2007, Kalton and Anderson, 1986; and Sudman, Sirken, and Cowan, 1988).

The first strategy for oversampling Korean and Vietnamese populations was geographic targeting using the same substrata used since 2003. These strata were created classifying exchanges based on the concentration of Korean and Vietnamese residing in the exchange[3] within selected counties. Under disproportionate stratified sampling, telephone numbers in exchanges located in areas with a relatively high proportion of members (high-density strata) were sampled at a higher rate than the numbers in the other areas (low-density strata). Since the stratification was based on information from the 2000 Census, we examined the observed sample from the previous cycles and reclassified the telephone exchanges using the sample distribution of these populations in previous cycles of CHIS. Reclassifying exchanges reflected changes since 2000 in the Korean and Vietnamese populations in these areas.

The second strategy to increase the number of Korean and Vietnamese interviews included supplemental samples from other frames (i.e., surname lists of the race-ethnic groups). This sampling

[3] Refer to the *CHIS 2003 Methodology Series: Report 1 Sample Design* for additional details on the creation of the substrata.

strategy is based on the concept of a dual frame design. In this approach, the landline sample is supplemented with a much less expensive sample drawn 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 is, 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 (landline 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 landline sample.

The identification of eligible (i.e., Korean or Vietnamese) adults in the list samples was done through a question in the screener interview. This strategy was relatively simple to implement and has good statistical properties, except for any measurement error that may be introduced by asking a question about the ethnicity of the adults at the beginning of the interview. Screening was not necessary for the cases sampled from the high/low density strata because these cases were part of the base landline 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 variance of estimates for these groups is not greatly reduced by this approach.

3. SAMPLING HOUSEHOLDS

This chapter describes the sample design and selection of households for CHIS 2009. We begin by defining the target population and the persons included in and excluded from 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 the selection of telephone numbers in order to achieve the stated goals. We also review the statistical issues considered in arriving at the allocation of the sample for the different components of the survey.

3.1 Population of Interest

As in previous CHIS cycles, the 2009 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 younger). Eligible residential households included houses, apartments, and mobile homes occupied by individuals, families, multiple families, extended families or multiple unrelated persons, if 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. These include those living in households without any telephone service, and children and adolescents living in a household without a parent or legal guardian.

3.2 Sample Design

The principal goals of the CHIS 2009 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 2009, a landline sample and surname list samples and a statewide cell-phone sample were drawn in order to meet these goals. The cell phone sample supplemented the landline sample. These samples are described in the following sections.

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.

3.2.1 Landline Sample

The CHIS 2009 sample had a goal of completing 41,900 statewide landline adult interviews form the base, surname, and geographic samples. The goal of the cell phone sample was 1,100 cell-only adult interviews. We also expected to get 1,420 adult interviews from adults with both telephone services (both landline and a cell phone) from the cell phone sample. This statewide goal includes 1,000 each of Koreans and Vietnamese from the landline, geographic, and surname samples. The statewide landline sample includes a base landline sample of 37,000 adult interviews and a supplemental geographic sample of 4,200 interviews.

The landline goal for adult interviews in CHIS 2009 was 4,700 interviews (10 percent) lower than the adjusted landline sample goal for CHIS 2007[4]. Although the number of child and adolescent interviews was not predetermined, we expected to get between 3,300 to 3,500 completed adolescent interviews (depending on compliance since parental consent and adolescent agreement are required) and between 8,800 to 9,000 child interviews from the landline/surname samples based on the CHIS 2007 results.

The goals of the base landline sample required allocating the sample into the sampling strata using a compromise between objectives. 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 individual, county-level estimates is to assign each county an equal

[4] Goals excluded the geographic and surname samples in CHIS 2005 and 2007

sample size.

The stratification of California's 58 counties used in CHIS 2009 was the same as the one used since 2005. The design consisted of 44 strata, with 41 single-county strata and 3 strata with multiple counties. The multiple-county strata were created by grouping the remaining counties into three geographic areas. The stratum assignment was based on the population residing in the county. Table A-1 in the appendix shows the assignment of counties to geographic strata across the CHIS cycles. Table 3-1 shows the 44 geographic sampling strata, and the targets for the base sample and geographic supplement samples of adult interviews for CHIS 2009.

Because of the need to produce reliable estimates at the county level, the sample allocation was not proportional to the population in the 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 target sample sizes from medium and smaller counties was fixed at 400 or 500 interviews. The remaining sample was allocated proportional to the population size. More details about the landline sample are given after discussing the designs for the other samples.

Table 3-1. Targeted number of completed adult interviews for the landline sample by county

	Stratum	Base	Geographic supplement	Total	Population size
	State Total	37,700	4,200	41,900	
1	Los Angeles	7,800	0	7,800	Over 9,000,000
2	San Diego	2,320	2,480	4,800	1,200,000 or greater
3	Orange	2,240	0	2,240	
4	Santa Clara	1,240	0	1,240	
5	San Bernardino	1,440	0	1,440	
6	Riverside	1,540	0	1,540	
7	Alameda	1,150	0	1,150	
8	Sacramento	1,150	0	1,150	
9	Contra Costa	850	0	850	800,000 to 1,200,000
10	Fresno	610	0	610	
11	San Francisco	760	0	760	500,000 to 800,000
12	Ventura	760	0	760	
13	San Mateo	570	0	570	
14	Kern	570	0	570	
15	San Joaquin	500	0	500	

Table 3-1. Targeted number of completed adult interviews for the landline sample by county (Continued)

	Stratum	Base	Geographic supplement	Total	Population size
16	Sonoma	500	0	500	Medium counties 100,000 to 500,000
17	Stanislaus	500	0	500	
18	Santa Barbara	500	0	500	
19	Solano	500	0	500	
20	Tulare	500	0	500	
21	Santa Cruz	500	0	500	
22	Marin	500	1,420	1,920	
23	San Luis Obispo	500	0	500	
24	Placer	500	0	500	
25	Merced	500	0	500	
26	Butte	500	0	500	
27	Shasta	500	0	500	
28	Yolo	500	0	500	
29	El Dorado	500	0	500	Medium counties 100,000 to 500,000
30	Imperial	500	0	500	
31	Napa	500	0	500	
32	Kings	500	0	500	
33	Madera	500	0	500	
34	Monterey	500	0	500	
35	Humboldt	500	300	800	
36	Nevada	500	0	500	Small counties less than 100,000 population per county
37	Mendocino	500	0	500	
38	Sutter	500	0	500	
39	Yuba	500	0	500	
40	Lake	500	0	500	
41	San Benito	500	0	500	
42	Colusa, Glenn, Tehama	400	0	400	Small counties combined
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	400	0	400	
44	Amador, Alpine, Calaveras, Inyo, Mariposa, Mono, Tuolumne	400	0	400	

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

3.2.2 Supplemental Geographic Samples

In CHIS 2009 supplemental geographic samples were added at the request of San Diego 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 landline sample for the specific county and were drawn in the same way as the base landline 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 landline sample and supplemental geographic sample instruments.

Table 3-1 also shows the targeted number of adults for the supplemental geographic samples in CHIS 2009. Combining the base landline and geographic samples the total targeted number of adults was 4,800 for San Diego County, 1,920 for Marin County, and 800 for Humboldt County.

3.2.3 Stratification of the Landline Sample

In this section, we describe the detailed steps used to select the sample of telephone numbers for the landline sample. These steps include stratifying the telephone numbers, selecting the sample of numbers after adjusting for expected losses due to nonresponse, and subsampling the numbers based on refusal status to improve the efficiency of the sample.

The first step was stratifying the sampling frame of 100-banks with one or more listed telephone numbers into the nonoverlapping strata, each corresponding to a county or a group of counties as shown in Table 3-1. The procedure for assigning the numbers to strata was the same as that used in previous CHIS cycles. The geographic information required for stratification was available only at the exchange level[5], 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.

To solve the stratification problem, the procedure used coverage reports for each county produced by MSG, the sampling vendor. The coverage reports listed all the exchanges in the county. For each exchange, the report showed the total number of listed households in the exchange and the

[5]A 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.

proportion of listed households that were within the county. After combining information from 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 then assigned to the county with the most listed households. There was also interest in obtaining a better sample distribution for Los Angeles County by Service Planning Areas (SPAs). 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 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). An analysis done in CHIS 2003 to help with the allocation 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. 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 2009. 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 from 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.

Table 3-2 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 3-2. Definition of sampling substratum, 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	27	52,541
	1.013	San Gabriel SPA	High	80	155,482
	1.014	Metro SPA	High	126	131,134
	1.017	South SPA	High	29	41,418
	1.018	South Bay SPA	High	50	83,817
	1.021	Antelope Valley SPA	Low	54	107,094
	1.022	San Fernando SPA	Low	424	688,546
	1.023	San Gabriel SPA	Low	255	390,607
	1.024	Metro SPA	Low	191	253,140
	1.025	West SPA	Low	257	322,077
	1.026	South SPA	Low	186	309,927
	1.027	East SPA	Low	174	341,564
	1.028	South Bay SPA	Low	264	432,528
2. San Diego	2.1	N/A	High	84	149,865
	2.2	N/A	High	551	896,379
3. Orange	3.01	N/A	High	278	354,707
	3.02	N/A	Low	392	609,897
4. Santa Clara	4.01	N/A	High	176	191,540
	4.02	N/A	Low	296	406,496
Total				3,894	5,918,759

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

3.2.4 Supplemental Surname List Samples

The second type of supplemental sample was the surname sample used to increase the number of completed interviews of adults of Korean and Vietnamese descent. The statewide goal was 1,000 completed adult interviews from each ethnic group from the combined landline and surname samples. As in previous cycles of CHIS, the Korean and Vietnamese supplemental samples were drawn from lists of telephone numbers with Korean and Vietnamese surnames maintained by the sampling vendor. We screened the telephone numbers in these samples to determine 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 2009. The targets of the surname list sample were adjusted during data collection, as the actual landline and surname samples yields became known. In CHIS 2009, the landline sample did not produce the expected number of Korean cases; therefore, we drew additional numbers from the list frame during the data collection period.

Table 3-3. Targeted number of completed adult interviews for the Korean and Vietnamese samples

	Targeted number of adult interviews		
Subgroup	Landline sample	Supplemental list sample	Total
Korean	308	692	1,000
Vietnamese	214	786	1,000
Total	522	1478	2,000

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

The list frames were created by the sampling vendor by compiling lists of surnames likely to be Korean or Vietnamese from telephone directories in California. The vendor provided four non-overlapping surname frames; the first two frames included telephone numbers whose associated surnames were very likely to be Korean *only* or Vietnamese *only*, and not any other ethnic group. The last two frames included those surnames likely to be either Korean or some other group, or Vietnamese and some other group.

Separate samples were drawn from each of the four frames. Households were eligible for the extended interview 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 2009 and the number of telephone numbers drawn from each frame.

Table 3-4. Surname frames and sample sizes

Surname frame	Number of records	Sample size
Korean only	38,977	12,355
Vietnamese only	103,712	32,925
Korean and some other race but Vietnamese	81,530	60,348
Vietnamese and some other race but Korean	26,630	21,087

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

3.2.5 Cell Phone Sample

The CHIS 2009 cell phone sample had a state-wide target of completing 1,100 interviews with adults living in households with only cellular telephone service, called cell phone-only households. Since all households reached through the cell sample were eligible in CHIS 2009, it was expected to complete about 1,420 adult interviews from adults with both telephone services (landline and a cell phone). In contrast with CHIS 2007, children and adolescents from the cell phone sample were also sampled and interviewed. The 2009 CHIS cell phone sample design was based on the results of the 2007 cell phone sample.

The cell phone sample design was different from the landline design and presented its own challenges. The main cell phone sample was drawn by the sampling vendor using the latest Telcordia database. This sample was selected from 1000-series blocks in California dedicated to wireless service.[6] Telephone numbers that were ported from a landline to a cell phone could not be selected from these exchanges because these numbers were in exchanges assigned to landlines. To address this problem, telephone numbers identified as ported cell phones in the base landline sample were included as part of the cell phone sample. The ported numbers were identified by disposition code in the CSS (see codes WR and CP in Table 2-1). There were close to 3,000 ported cell phone numbers (0.3 percent) identified in the landline sample. This is twice the percentage of ported numbers identified in 2007. The remainder of this section discusses the sampling of the main cell sample.

One problem that is unique to cell phone sampling is assigning a geographic location to a number. Although cell phone numbers are sampled from exchanges assigned to wireless service, the geographic area covered by the exchange does not necessarily indicate where the owner of the number resides. This is because the cell phone exchange generally corresponds to where the cell phone was purchased. Thus, the cell sample could not be stratified to match the stratification of the landline sample. Instead, the objective was to devise a sample design that would yield approximately equal numbers of interviews to 7 geographic regions in California.

Since there was not exact information on the geographic area covered by the cell phone exchange, the sampling strata were created and assigned to regions in an indirect way. First, we determined the counties covered by the area code using the number of households in each area code.

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

While some area codes were completely contained in a single county (Los Angeles County, for example); most area codes covered multiple counties. Each area code was assigned to the county with the greatest proportion of households in the area code, and thus to the region containing that county. For example, 43 percent of all the households in area code 209 fell in San Joaquin County, the highest county percentage. As a result, area code '209' was assigned to San Joaquin County and to the San Joaquin Valley region. Table 3-5 shows the area codes in California, their corresponding assigned county and the percentage of the households in the area code that fall into that county.

Table 3-5. Assignment of cell phone area codes to counties and regions

Mapped Region Number	California Area Code	Assigned County	(%) in Assigned County
1 - Northern & Sierra Counties	530	Butte	15
2 - Greater Bay Area	408	Santa Clara	100
	415	San Francisco	78
	510	Alameda	91
	650	San Mateo	84
	707	Sonoma	32
	925	Contra Costa	76
3 - Sacramento Area	916	Sacramento	81
4 - San Joaquin Valley	209	San Joaquin	41
	559	Fresno	58
	661	Kern	58
5 - Central Coast	805	Ventura	55
	831	Monterey	52
6 - Los Angeles	213	Los Angeles	100
	310	Los Angeles	100
	323	Los Angeles	100
	562	Los Angeles	85
	626	Los Angeles	100
	818	Los Angeles	100
7 - Other Southern California	619	San Diego	100
	714	Orange	100
	760	San Diego	34
	858	San Diego	100
	909	San Bernardino	72
	949	Orange	100
	951	Riverside	100

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Unlike the landline sample where the numbers were drawn from banks with a 100 numbers,

the cell phone numbers were drawn from groups of 1,000 numbers. Another difference between the two samples is the lack of detailed demographic and socio-economic information (e.g. number of households, percentage of homeowners, African Americans, etc.) on the geographic area from which the cell phone is sampled.

When determining the sample size to draw, we used the observed response rates and proportion of cell-only households within regions from the cell sample in 2007. Furthermore, the misclassification rate between the sampled region and the self–reported region observed in 2007 was used to allocate the sample. For example, only 47 percent of the cases sampled in Region 1 reported to be located in Region 1; while close to 37 percent of the cases sampled in the same region were located in Region 3. When the sample was allocated, we expected that 37 percent of the cases sampled in Region 1 would report to be in Region 3. As a result, we increased the sample drawn in Region 1 to compensate for the cases to be lost to Region 3. The sample drawn in Region 3 was reduced because we expected additional cases from Region 1. The sample was allocated numerically changing the number of cases to draw in each region and area code so that the expected number of self-reported cases was the same across regions (i.e., same sample by self-reported region) with a total sample of 1,100 cell-only adult interviews, while minimizing the design effect.

Table 3-6 shows the target yield along with the sample size drawn from each region. Although the sampling rate assignment was done at the region level, the sample was selected using the exchange as a sampling stratum (see Table A-6 in Appendix A).

Table 3-6. Cell sample sampling size and targets by geographic regions

California Region	Sampled Phone Numbers	Sampled Yield	Self-Reported Yield
1 - Northern & Sierra Counties	13,200	263	157
2 - Greater Bay Area	8,700	160	157
3 - Sacramento Area	3,300	63	157
4 - San Joaquin Valley	9,700	182	157
5 - Central Coast	7,400	151	157
6 - Los Angeles	10,200	166	158
7 - Other Southern California	6400	114	157
Total	58,900	1,100	1,100

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

There were also differences in the way the cell sample was processed after it was selected.

Unproductive numbers in the cell phone sample (i.e., nonworking and business telephone numbers) could not be purged using directory matching because no cell phone directories exist. In addition, there are prohibitions on predictive dialing of cell phone numbers, so the other components of the CSS purging for nonproductive or nonworking numbers could not be done. Thus, the full selected cell phone sample was sent to be dialed by interviewers.

3.3 Sample Selection

The number of telephone numbers selected in any telephone 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 are not residential. For CHIS 2009 the sample of telephone numbers was inflated to deal with losses due to the following sources:

- Nonworking, nonresidential, and never answered numbers;
- Nonresponse to the screening interview;
- Nonresponse to the extended interview; and
- Ineligible households in the surname list and cell phone samples.

The first three sources noted above are typical of all telephone surveys. To deal with these losses we used information from CHIS 2007 to estimate the percentage of 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. Estimates of the eligibility rates were taken from the CHIS 2007 surname samples.

Taking all of these factors into consideration, 1,034,395 telephone numbers[7] were sampled for CHIS 2009. Not all the telephone numbers were selected at the same time, as the sample design was modified several times during the field period to reflect the observed yield and changes in the targeted number of completed interviews. After each selection, duplicate telephone numbers (those numbers that had been previously sampled) were removed from the samples. Table 3-7 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. The data collection procedures are discussed in *CHIS 2009 Methodology Series: Report 2 - Data*

[7] This total excludes the area sample matched numbers.

Collection Methods.

Table 3-7. Number of telephone numbers drawn by sample type

Sample type	Number of telephone numbers drawn
Base landline sample	756,425
Geographic supplemental sample	88,481
Surname List samples	
Korean only	12,355
Korean and other	32,925
Vietnamese only	60,348
Vietnamese and other	21,087
Cell phone sample	62,774
Total	1,034,395

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

3.4 Expected Design Effect

Previous sections described the allocation of the sample of telephone numbers by sampling stratum and substratum and noted that it involved compromises 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 under different allocations. These methods helped guide the allocation of the CHIS 2009 sample.

If CHIS 2009 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-α)·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 confidence interval is not appropriate for CHIS 2009 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 2009 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 from proportionate allocation across 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 can be computed as

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

where W_h is the proportion of stratum *h* in the population computed as $W_h = N_h \left(\sum N_h \right)^{-1}$, 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 a 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

$$n_{eff} = \frac{n}{D}. \tag{3}$$

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

In CHIS 2009, we expected to complete 41,900 adult interviews from the landline sample

(the supplemental geographic sample 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 3-8. 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 and the effect of sampling from the surname frames.

Table 3-8. Expected design effects and effective adult sample size associated with the sample allocation for the base landline sample

Race* and ethnicity	Expected nominal sample size	Expected design effect	Expected effective sample size
Total	41,900	1.27	33,096
White alone non-Latino	21,979	1.35	16,311
African American alone non-Latino	2,072	1.13	1,839
American Indian alone non-Latino	383	1.51	254
Asian alone non-Latino	4,292	1.12	3,848
Native Hawaiian or Pacific Islander	142	1.16	122
Multiple race non-Latino	660	1.28	515
Latino	12,373	1.20	10,311

* Office of Management and Budget definition of race

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

For example, the expected yield from the CHIS 2009 sample for Latino was 12,373 adults for the landline sample. Due to the allocation of the sample, the expected effective sample size was 10,311. 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

$$\left(0.5-1.96\sqrt{\frac{0.5^2}{254}},0.5+1.96\sqrt{\frac{0.5^2}{254}}\right)=(0.4385,0.5615)$$

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 was selected, interviewers called the numbers and conducted interviews with sampled persons within the household. This chapter describes the procedures for selecting the sample of persons within households for CHIS 2009. 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 a given 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 2009.

4.1 Sampling Alternatives

The general idea for the sample design over the CHIS cycles has been to sample one adult randomly from all the adults in the 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 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 solve this analytic problem, a second sampling alternative was adopted and has been used

since CHIS 2001. 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 eligible adults in the household. In the second phase, a child is sampled from all eligible children associated with the sampled adult. Since the sampling of children is a two-phase procedure, the probability of selection of the child is the product of the probability of selecting the adult (phase one) and the probability of selecting the child from all children associated with that adult (phase two). Adolescents are sampled in the same way, that is, one adolescent is selected 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 2009. 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 does 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 interviewed. Therefore, unassociated children and adolescents in a household could not be randomly linked to an adult in the household in 2009 and most previous cycles of the survey. The bias due to excluding unassociated children and adolescents was expected to be very small; however, it is not possible to evaluate it.

4.2 Child First Procedure

In the first two 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, the child and adolescent interviews could 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 2001 and 2003, 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 associated with the sampled adult, 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 conducted without waiting for the completion of the adult interview. When the child-first criteria were not met, the sampling for children or adolescents was not done until the adult was interviewed. For the first time in 2009, children and adolescents were selected in the cell sample; however, the child-first procedure was not used for the cell sample.

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 2009. This table was produced using the variables created during the CATI interview and they may not match the final counts of households with children and adolescents at the different stages. In CHIS 2009, 21,151 households with children[8] completed the screener interview. The child-first procedure was used in 28.1 percent of households with children with a completed screener interview. A child interview was completed in 63.2 percent (3,758 interviews) of the households with children where the child-first procedure was used. In comparison, a child interview was completed in only 30.4 percent (4,620 cases) of households with children where the procedure was not used.

Although the child-first procedure was intended to increase the number of child interviews, it did not have a large effect on the number of adolescent interviews. The child-first procedure was used in 12.7 percent of the households with adolescents and 32.1 percent of those completed the adolescent interview. In only 16.8 percent of the households with adolescents where the child first procedure was not used the adolescent interview was completed.

See *CHIS 2009 Methodology Series: Report 2 - Data Collection* for more detail on the child-first procedures and further evaluation of the yields.

[8] This number is includes households with children but where the sample adult is not related to any of the children in the household.

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

Households with children that completed a screener interview	Count	Percentage
Total number of households with children	21,151	100.0
Participated in the child-first procedure	5,949	28.1
Did not participate in the child-first procedure	15,202	71.9

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	5,949	100.0
Completed the extended interview	3,758	63.2
Did not complete the extended interview	2,191	36.8

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	15,202	100.0
Completed the extended interview	4,620	30.4
Did not complete the extended interview	10,582	69.6

Households with adolescents that completed a screener interview	Count	Percentage
Total number of households with adolescents	16,516	100.0
Participated in the child-first procedure	2,090	12.7
Did not participate in the child-first procedure	14,426	87.3

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	2,090	100.0
Completed the extended interview	671	32.1
Did not complete the extended interview	1,419	67.9

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	14,426	100.0
Completed the extended interview	2,422	16.8
Did not complete the extended interview	12,004	83.2

*Includes landline and surname samples

Source: UCLA Center for Health Policy Research, 2009 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 2009 for the landline sample and the area sample was the same as that used since 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 $N = 1, 2, 3, \ldots$;
- 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 or the screener respondent does not know the birthdays of the other adults, 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.

4.3.1 Adult Sampling in the Cell Sample

Procedures for the sampling of adults within the cell-only household were developed and implemented in the CHIS 2005 cell-phone pilot and were based on principles similar to those used in landline RDD surveys (Brick, Edwards, and Lee 2007). The same procedure was used in CHIS 2009. In this approach adults were sampled during the screening interview.

In cell-only households with only one adult, no sampling was required. In households with more than one adult, sampling adults depended on whether other household members shared the cell phone. If adults shared the cell phone, the same within-household sampling method used in base landline sample was implemented. That is the screener respondent (SR) is randomly selected for the adult interview with a probability equal to the inverse of the number of adults in the household. In case the SR is not selected, then one adult other than the SR is selected for the adult interview using the next birthday method. If the cell phone was not shared, then the SR is sampled.

This sampling scheme assumes that, in cell-only households with more than one adult, each adult has a cell phone (or shares a different cell phone) if the sampled cell phone is not shared. This weakness in the sampling scheme was recognized during the pilot in 2005. However, the alternative approach required asking the full battery of items required to ascertain the cell phone status of each adult in the household, which was viewed as a heavy burden that could detract from gaining cooperation. In other words, while the sampling scheme did not address all possible forms of within-household undercoverage, it was believed this was a good compromise between reducing the potential for increased nonresponse and coverage errors.

4.4 Child Sampling

In an earlier cycle of CHIS, the child sampling procedure was modified 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). The same sampling procedure was used in CHIS 2009 in the landline and surname samples. Previously all children were sampled at the same rate. If there were only younger or older children in the sampled households, a child was selected with equal probability of selection. 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

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

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 is 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 2009. Children were selected with unequal probability of selection in approximately 25 percent of the households with children.

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 only 0 to 5 years old	13,614	45.2
	Household with children only 6 to 11 years old	8,979	29.8
Unequal probability	Household with children 0 to 5 and 6 to 11 years old	7,537	25.0
Total		30,130	100.0

* Includes all samples

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

4.5 Adolescent Sampling

The sampling method used in CHIS 2009 to select an adolescent did not change from previous cycles. That is, an adolescent was sampled with equal probability from among all eligible adolescents associated with the sampled adult in a household. In the landline and list samples, 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 otherwise. 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.

Adolescents were also sampled for the first time in the cell phone sample in CHIS 2009. Similar to the child selection, adolescents were selected in section G of the adult extended interview and no child-first procedures were used in the cell phone sample. *CHIS 2009 Methodology Series*: *Report 5 - Weighting and Estimation* describes how the probabilities of selection are computed for the sampled adults, children, and adolescents in the landline, surname, and cell phone samples.

5. ACHIEVED SAMPLE SIZES

This chapter summarizes the number of completed interviews in CHIS 2009 for the landline, surname and cell phone samples and the relationship between the targeted and the achieved numbers. As mentioned in the previous chapters, the targeted goals for CHIS 2009 were stated in terms of the total number of completed adult interviews. 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 nonresponse. These reasons were discussed in more detail in Chapter 3. Detailed information about the response rates is presented in *CHIS 2009 Methodology Series: Report 4 – Response Rates.*

Table 5-1 shows the number of completed interviews by sample type compared to the adjusted targets. The table shows that, in general, the target goals were met in CHIS 2009 at the state level.

Table 5-1. Number of completed interviews by type of sample

Sample type	Number of Completed interviews	Goal	% Completed
Landline, and geographic samples			
Adults	42,682	41,900	101.9
Child	7,918	—	—
Adolescent	3,002	—	—
Surname samples			
Adults	1,885	—	—
Child	545	—	—
Adolescent	178	—	—
Cell phone sample			
Cell phone only			
Adults	1,187	1,100	107.9
Child	201	—	—
Adolescent	52	—	—
Both cell phone and landline			
Adults	1,860		
Child	281	—	—
Adolescent	147	—	—
Total	59,938		

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Table 5-2 also shows the number of completed interviews as percentages of the targeted number of adult interviews for the landline 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. The targets were met or surpassed in 29 strata of the 44 strata based on the sampling location information that was available at the time of data collection. In the same number of strata the target number of completes was met or surpassed based on the self reported location.

Table 5-2. Number of completed adult interviews for the base landline, surname and geographic samples by sampling and self-reported stratum*

	Sampling location		Self-reported location	
Stratum	Completed interviews	% of Targeted interviews	Completed interviews	% of Targeted interviews
State	44,567	106.4	44,567	106.4
Los Angeles	8,710	111.7	8,731	111.9
San Diego	5,014	104.5	5,012	104.4
Orange	2,636	117.7	2,587	115.5
Santa Clara	1,607	129.6	1,659	133.8
San Bernardino	1,460	101.4	1,456	101.1
Riverside	1,609	104.5	1,631	105.9
Alameda	1,191	103.6	1,156	100.5
Sacramento	1,226	106.6	1,230	107.0
Contra Costa	895	105.3	946	111.3
Fresno	667	109.3	663	108.7
San Francisco	757	99.6	745	98.0
Ventura	898	118.2	920	121.1
San Mateo	590	103.5	561	98.4
Kern	577	101.2	576	101.1
San Joaquin	517	103.4	515	103.0
Sonoma	512	102.4	523	104.6
Stanislaus	474	94.8	452	90.4
Santa Barbara	613	122.6	602	120.4
Solano	480	96.0	478	95.6
Tulare	473	94.6	480	96.0
Santa Cruz	503	100.6	484	96.8
Marin	2,048	106.7	2,037	106.1
San Luis Obispo	478	95.6	484	96.8
Placer	502	100.4	501	100.2
Merced	493	98.6	512	102.4
Butte	493	98.6	508	101.6
Shasta	502	100.4	526	105.2
Yolo	524	104.8	511	102.2
El Dorado	494	98.8	514	102.8

Table 5-2. Number of completed adult interviews for the base landline, surname and geographic samples by sampling and self-reported stratum* (Continued)

	Sampling stratum		Self-reported stratum	
	Completed interviews	% of Targeted interviews	Completed interviews	% of Targeted interviews
Imperial	539	107.8	532	106.4
Napa	485	97.0	495	99.0
Kings	478	95.6	477	95.4
Madera	535	107.0	530	106.0
Monterey	426	85.2	490	98.0
Humboldt	844	105.5	846	105.8
Nevada	537	107.4	528	105.6
Mendocino	600	120.0	580	116.0
Sutter	468	93.6	475	95.0
Yuba	466	93.2	426	85.2
Lake	525	105.0	520	104.0
San Benito	548	109.6	502	100.4
Colusa, Glenn, Tehama	382	95.5	356	89.0
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	403	100.8	422	105.5
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	388	97.0	388	97.0

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

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Table 5-3 shows the number of completed child and adolescent interviews for the landline, geograpghic, and surname samples. Because there were not predetermined targets by stratum for children and adolescents, columns for the percentages of the targeted number of interviews are not included in the table. However, we expected to get between 8,800 and 9,000 child interviews based on the CHIS 2007 data and 8,463 interviews were completed. Similarly, we expected between 3,200 and 3,500 completed adolescent interviews and 3,180 were completed.

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

Stratum	Completed interviews			
	Child		Adolescents	
	Sampling location	Self-reported location	Sampling location	Self-reported location
State Total	8,463	8,463	3,180	3,180
Los Angeles	1,703	1,697	636	641
San Diego	932	936	339	341
Orange	530	524	182	174
Santa Clara	417	423	120	124
San Bernardino	305	308	108	108
Riverside	293	300	123	123
Alameda	264	248	98	98
Sacramento	222	220	72	71
Contra Costa	145	163	56	56
Fresno	153	149	57	58
San Francisco	99	98	21	23
Ventura	166	172	68	70
San Mateo	109	101	37	33
Kern	136	133	53	51
San Joaquin	113	113	46	46
Sonoma	96	97	40	41
Stanislaus	101	95	40	38
Santa Barbara	94	91	43	43
Solano	76	78	39	41
Tulare	102	106	42	43
Santa Cruz	102	96	38	36
Marin	325	323	127	125
San Luis Obispo	60	60	25	25
Placer	105	108	30	32
Merced	110	116	47	49
Butte	67	72	29	32
Shasta	73	77	23	25
Yolo	116	113	58	56

Table 5-3. Number of completed child and adolescent interviews for base landline, surname and geographic samples by sampling and self-reported stratum(Continued)

	Completed interviews			
	Child		Adolescents	
Stratum	Sampling location	Self-reported location	Sampling location	Self-reported location
El Dorado	75	77	32	33
Imperial	138	137	51	51
Napa	66	67	22	22
Kings	122	121	41	41
Madera	136	137	44	43
Monterey	91	107	35	38
Humboldt	110	112	46	46
Nevada	73	70	30	27
Mendocino	79	77	36	35
Sutter	86	88	27	30
Yuba	97	90	45	40
Lake	77	76	42	41
San Benito	139	130	57	55
Colusa, Glenn, Tehama	67	63	34	32
Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	50	51	18	19
Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	43	43	23	24

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Table 5-4 shows the number of completed adult interviews by ethnicity and sample type. The supplemental sample targets were revised during the data collection period as experience was gained on the actual landline sample yield. The target was exceeded for the number of completed Vietnamese adult interviews but was not met for the Korean interviews.

Table 5-4. Number of completed adult interviews by ethnicity and sample type

	Number of completed interviews		
Sample	Korean only	Vietnamese only	Other*
Landline sample	288	226	42,168
Korean only list	336	1	5
Korean and other list	279	12	7
Vietnamese only list	5	1008	57
Vietnamese and other list	3	162	10
Total	911	1409	42,247
Target	1,000	1,000	N/A
Percentage of Target	91.10	140.90	N/A

* Korean or Vietnamese and other group.

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Table 5-5 shows the number of adult completed interviews in cell-only households by type of cell phone (i.e., ported landline number and exchanges assigned for wireless service). The difference between the sampled region and the self-reported region is large, as expected, for the cell phone numbers sampled from the exchanges assigned to wireless services. The goal of 1,100 completed cell-only adult interviews was exceeded. Similarly, Table 5-6 shows the number of completed adult interviews with both telephone services from the cell sample. The anticipated yield of 1,420 interviews of adults with landline and cell phone was also exceeded for this sample.

Table 5-5. Number of completed adult interviews for the cell phone sample by sampling and self-reported region geographic regions

	Ported		Wireless assigned		Total	
	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum
Cell only households	75	75	1,112	1,112	1,187	1,187
1 - Northern & Sierra Counties	8	9	235	172	243	181
2 - Greater Bay Area	16	16	170	173	186	189
3 - Sacramento Area	7	6	73	148	80	154
4 - San Joaquin Valley	2	2	204	185	206	187
5 - Central Coast	5	5	145	140	150	145
6 - Los Angeles	17	14	166	159	183	173
7 - Other Southern California	20	23	119	135	139	158

Table 5-5. Number of completed adult interviews for the cell phone sample by sampling and self-reported region geographic regions (Continued)

	Ported		Wireless assigned		Total	
	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum
Cell phone and landline households	102	102	1,758	1,758	1,860	1,860
1 - Northern & Sierra Counties	3	4	423	346	426	350
2 - Greater Bay Area	32	30	272	242	304	272
3 - Sacramento Area	5	5	121	236	126	241
4 - San Joaquin Valley	4	5	315	295	319	300
5 - Central Coast	9	9	239	230	248	239
6 - Los Angeles	23	23	210	221	233	244
7 - Other Southern California	26	26	178	188	204	214
Total	177	177	2,870	2,870	3,047	3,047

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Tables 5-6 and 5-7 shows the number of completed child and adolescent interviews in cell-only households by type of cell phone (i.e., ported landline number and exchanges assigned for wireless service) in cell only households and households with both telephone services. There was no goal set for these type of interviews.

Table 5-6. Number of completed child interviews for the cell phone sample by sampling and self-reported region geographic regions

	Ported		Wireless assigned		Total	
Child interviews	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum
Cell only households	9	9	192	192	201	201
1 - Northern & Sierra Counties	1	1	37	34	38	35
2 - Greater Bay Area	2	2	24	24	26	26
3 - Sacramento Area	1	1	10	14	11	15
4 - San Joaquin Valley	1	1	51	49	52	50
5 - Central Coast	0	0	27	25	27	25
6 - Los Angeles	1	1	22	24	23	25
7 - Other Southern California	3	3	21	22	24	25

Table 5-6. Number of completed child interviews for the cell phone sample by sampling and self-reported region geographic regions (Continued)

Child interviews	Ported		Wireless assigned		Total	
	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum
Cell phone and landline households	11	11	270	270	281	281
1 - Northern & Sierra Counties	0	0	53	49	53	49
2 - Greater Bay Area	4	5	39	38	43	43
3 - Sacramento Area	0	0	27	36	27	36
4 - San Joaquin Valley	0	0	57	55	57	55
5 - Central Coast	2	2	35	34	37	36
6 - Los Angeles	3	2	30	31	33	33
7 - Other Southern California	2	2	29	27	31	29
Total	20	20	462	462	482	482

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Table 5-7. Number of completed adolescent interviews for the cell phone sample by sampling and self-reported region geographic regions

Adolescent interviews	Ported		Wireless assigned		Total	
	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum	Sampling stratum	Self-reported stratum
Cell only households	3	3	49	49	52	52
1 - Northern & Sierra Counties	1	1	15	10	16	11
2 - Greater Bay Area	0	0	5	6	5	6
3 - Sacramento Area	0	0	3	7	3	7
4 - San Joaquin Valley	0	0	11	11	11	11
5 - Central Coast	0	0	6	5	6	5
6 - Los Angeles	2	1	6	5	8	6
7 - Other Southern California	0	1	3	5	3	6
Cell phone and landline households	8	8	139	139	147	147
1 - Northern & Sierra Counties	0	0	36	31	36	31
2 - Greater Bay Area	6	6	29	25	35	31
3 - Sacramento Area	0	0	11	19	11	19
4 - San Joaquin Valley	0	0	23	24	23	24
5 - Central Coast	0	0	18	19	18	19
6 - Los Angeles	2	2	9	10	11	12
7 - Other Southern California	0	0	13	11	13	11
Total	11	11	188	188	199	199

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

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

REFERENCES

Blumberg SJ, Luke JV.(2010). Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2010. National Center for Health Statistics. Available from: http://www.cdc.gov/nchs/nhis.htm. February 10, 2008.

Brick, J.M., and Waksberg, J. (1991). Avoiding sequential sampling with RDD. *Survey Methodology*, 17(1), 27-41.

Brick, J.M., Edwards. W.S., Lee S. (2007). Sampling telephone numbers and adults, interview length, and weighting in the California Health Interview Survey cell phone pilot study. Public Opinion Quarterly, 71:793-813.

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., Waksberg, J., Kulp, D., and Starer, A. (1995). Bias in list-assisted telephone surveys. *Public Opinion Quarterly*, 59(2), 218-235.

Delnevo C. D., Gundersen D. A., and Hagman B. T. (2008). Declining estimated prevalence of alcohol drinking and smoking among young adults nationally: artifacts of sample undercoverage?, American Journal of Epidemiolofy,167(1):15-9.Department of Health Education and Behavioral Sciences, School of Public Health, University of Medicine and Dentistry of New Jersey, New Brunswick, NJ 08901, USA.

Edwards, W.S., Brick, J.M., Flores Cervantes, I., DiSogra, C., and Yen, W. (2002). Sampling race and ethnic groups in RDD surveys. *Proceedings of the Survey Methods Section of the American Statistical Association* [CD-ROM].

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.

Fahimi, M., Kulp, D., and Brick, J. M. (2008). "Bias in List-Assisted 100-Series RDD Sampling." Survey Practice, September 28, 2008.

Flores Cervantes, I, and Kalton, G. (2007), "Methods for Sampling Rare Populations in Telephone Surveys" in Lepkowski, J.M, Tucker, C., Brick, J. M., de Leeuw, E. D., Japec, L., Lavrakas, P. J., Link, M.W., and Sangster, R.L. (Eds.). *Advances in Telephone Survey Methodology*. New York. John Wiley & Sons.

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

Lee, S., Kurata, J., Nguyen, H.A., and Jawad, M. (2008). Linguistic minorities and nonresponse error."*Public Opinion Quarterly*, 72, 470-486.

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., and Meekins, B. (2007). Household Telephone Service and Usage Patterns in the U.S. in 2004: Implications for Telephone Samples, *Public Opinion Quarterly*, 71, 3-22.

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

APPENDIX A

Table A-1. Stratum definitions for CHIS 2001, 2003, 2005, 2007, and 2009

County	2005 , 2007, and 2009 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	
Lake	40	37
Mendocino	37	
Sutter	38	39
Yuba	39	

Table A-1. Stratum definitions for CHIS 2001, 2003, 2005, 2007, and 2009 (continued)

County	2005 , 2007, and 2009 Stratum	2001 and 2003 Stratum
Colusa Glen Tehama	42	38
Humboldt	35	35
Del Norte, Lassen Modoc Siskiyou Trinity	43	36
Plumas Sierra		40
Nevada	36	
Alpine Amador Calaveras Inyo Mariposa Mono Tuolumne	44	41

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

				Supplemental Samples				
				Surname**				
	Stratum	Base Sample Landline *	Cell Phone	Landline Geographic	Korean Only	Korean and Other	Vietnamese Only	Vietnamese/ Other	Total
	State	848,780	58,900	0	12,355	32,925	60,348	21,087	1,034,395
1	Los Angeles	213,781	10,200	0	4,958	11,509	11,660	5,384	257,358
2	San Diego	111,465	2,400	57,452	500	986	3,783	804	120,048
3	Orange	53,575	2,000	0	1,616	2,687	15,065	2,812	77,746
4	Santa Clara	27,555	1,200	0	844	3,545	11,070	2,907	47,090
5	San Bernardino	24,710	1,200	0	398	734	1,447	457	28,931
6	Riverside	26,461	800	0	360	504	1,584	375	30,035
7	Alameda	24,043	600	0	721	3,448	3,880	2,198	34,877
8	Sacramento	19,850	3,300	0	303	893	2,646	836	27,874
9	Contra Costa	16,587	1,200	0	266	843	943	514	20,366
10	Fresno	11,400	5,700	0	160	325	598	143	18,374
11	San Francisco	27,269	1,200	0	565	4,082	2,325	2,550	38,249
12	Ventura	15,972	0	0	137	272	509	147	16,964
13	San Mateo	13,266	600	0	331	1,415	583	812	17,040
14	Kern	9,100	500	0	85	87	265	49	10,201
15	San Joaquin	7,999	3,500	0	106	216	841	327	13,031
16	Sonoma	7,738	3,900	0	101	109	321	84	12,343
17	Stanislaus	7,170	0	0	61	71	232	45	7,563
18	Santa Barbara	11,476	3,600	0	64	98	191	49	15,562
19	Solano	9,214	0	0	55	112	286	85	9,716
20	Tulare	8,749	0	0	26	47	98	16	8,912
21	Santa Cruz	7,502	0	0	42	70	106	31	7,707
22	Marin	37,763	0	27,738	66	114	222	58	37,965
23	San Luis Obispo	6,230	0	0	39	36	116	29	6,439
24	Placer	7,288	0	0	62	89	209	44	7,648

Table A-2. Number of telephone numbers and addresses drawn by sample frame and sampling stratum (Comtinued)

				Supplemental Samples					
					Surname**				
	Stratum	Base Sample Landline *	Cell Phone	Landline Geographic	Korean Only	Korean and Other	Vietnamese Only	Vietnamese/ Other	Total
25	Merced	6,700	0	0	32	45	91	16	6,864
26	Butte	5,550	13,200	0	36	36	128	36	19,110
27	Shasta	5,610	0	0	31	14	55	8	5,700
28	Yolo	7,726	0	0	54	170	189	84	8,195
29	El Dorado	7,140	0	0	31	39	78	20	7,290
30	Imperial	9,521	0	0	11	22	22	5	9,566
31	Napa	7,960	0	0	19	17	41	13	8,034
32	Kings	7,307	0	0	15	12	39	3	7,367
33	Madera	8,019	0	0	9	14	33	4	8,064
34	Monterey	8,258	3,800	0	84	121	201	51	12,569
35	Humboldt	8,785	0	3,291	19	13	47	7	8,863
36	Nevada	6,671	0	0	23	20	49	8	6,754
37	Mendocino	7,467	0	0	10	16	45	9	7,542
38	Sutter	6,450	0	0	14	15	37	24	6,533
39	Yuba	7,257	0	0	12	20	34	13	7,322
40	Lake	7,866	0	0	13	7	18	2	7,883
41	San Benito	9,551	0	0	3	5	15	3	9,565
42	Colusa, Glenn, Tehama	3,950	0	0	14	9	33	8	4,004
43	Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	5,809	0	0	21	12	85	5	5,915
44	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	5,020	0	0	38	26	128	12	5,216

* Ported cell phone numbers identified in the landline sample are counted as part of the landline sample

** Not drawn by sampling stratum.

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

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

Stratum	Base Sample Landline	Cell Phone	Supplemental Samples					
				Surname**				
			Landline Geographic	Korean Only	Korean and Other	Vietnamese Only	Vietnamese/ Other	Total
State	42,682	3,047	0	342	298	1,070	175	47,614
1 Los Angeles	8,248	417	0	164	138	155	26	9,148
2 San Diego	2,378	110	2,543	8	11	62	10	5,122
3 Orange	2,097	110	0	69	49	314	58	2,697
4 Santa Clara	1,277	72	0	17	30	294	41	1,731
5 San Bernardino	1,414	59	0	8	8	23	3	1,515
6 Riverside	1,601	88	0	2	9	16	3	1,719
7 Alameda	1,050	65	0	19	15	66	6	1,221
8 Sacramento	1,167	173	0	9	1	46	7	1,403
9 Contra Costa	916	62	0	8	6	11	5	1,008
10 Fresno	656	166	0	1	1	4	1	829
11 San Francisco	704	64	0	7	7	24	3	809
12 Ventura	902	87	0	5	5	7	1	1,007
13 San Mateo	545	25	0	7	4	5	0	586
14 Kern	570	25	0	1	3	1	1	601
15 San Joaquin	496	73	0	2	1	13	3	588
16 Sonoma	515	60	0	1	2	3	2	583
17 Stanislaus	447	63	0	0	1	3	1	515
18 Santa Barbara	597	57	0	1	1	3	0	659
19 Solano	470	68	0	3	1	3	1	546
20 Tulare	480	89	0	0	0	0	0	569
21 Santa Cruz	483	75	0	0	0	1	0	559
22 Marin	528	26	1,500	1	1	5	2	2,063
23 San Luis Obispo	483	33	0	1	0	0	0	517
24 Placer	495	69	0	2	1	2	1	570
25 Merced	511	25	0	0	0	1	0	537

Table A-3. Number of adult completed interviews by sample type and self-reported stratum (Continued)

Stratum	Base Sample Landline	Cell Phone	Supplemental Samples					Total
				Surname**				
			Landline Geographic	Korean Only	Korean and Other	Vietnamese Only	Vietnamese/ Other	
26 Butte	508	118	0	0	0	0	0	626
27 Shasta	525	76	0	1	0	0	0	602
28 Yolo	504	106	0	1	1	5	0	617
29 El Dorado	512	47	0	1	0	1	0	561
30 Imperial	531	5	0	0	1	0	0	537
31 Napa	495	19	0	0	0	0	0	514
32 Kings	475	20	0	0	1	1	0	497
33 Madera	530	26	0	0	0	0	0	556
34 Monterey	486	108	0	3	0	1	0	598
35 Humboldt	529	30	317	0	0	0	0	876
36 Nevada	528	53	0	0	0	0	0	581
37 Mendocino	580	13	0	0	0	0	0	593
38 Sutter	475	54	0	0	0	0	0	529
39 Yuba	426	29	0	0	0	0	0	455
40 Lake	520	5	0	0	0	0	0	525
41 San Benito	502	24	0	0	0	0	0	526
42 Colusa, Glenn, Tehama	356	64	0	0	0	0	0	420
43 Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	422	76	0	0	0	0	0	498
44 Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	388	13	0	0	0	0	0	401

* Not drawn by sampling stratum.

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

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

				Supplemental Samples				
				Surname*				
Stratum	Base Sample Landline	Cell Phone	Landline Geographic	Korean Only	Korean and Other	Vietnamese Only	Vietnamese/ Other	Total
State	7,918	482	0	83	52	361	49	8,945
1 Los Angeles	1,596	58	0	26	21	50	4	1,755
2 San Diego	438	13	469	3	4	19	3	949
3 Orange	381	14	0	20	9	99	15	538
4 Santa Clara	291	11	0	8	6	103	15	434
5 San Bernardino	289	13	0	2	2	14	1	321
6 Riverside	291	12	0	2	0	5	2	312
7 Alameda	211	11	0	5	5	24	3	259
8 Sacramento	202	27	0	2	0	16	0	247
9 Contra Costa	152	10	0	6	1	2	2	173
10 Fresno	149	43	0	0	0	0	0	192
11 San Francisco	90	12	0	1	1	5	1	110
12 Ventura	168	17	0	1	0	3	0	189
13 San Mateo	94	5	0	3	1	3	0	106
14 Kern	131	9	0	0	0	2	0	142
15 San Joaquin	108	14	0	0	0	4	1	127
16 Sonoma	93	7	0	0	0	3	1	104
17 Stanislaus	94	10	0	0	0	1	0	105
18 Santa Barbara	91	6	0	0	0	0	0	97
19 Solano	74	10	0	1	0	2	1	88
20 Tulare	106	17	0	0	0	0	0	123
21 Santa Cruz	96	6	0	0	0	0	0	102
22 Marin	84	3	238	0	0	1	0	326
23 San Luis Obispo	59	4	0	1	0	0	0	64
24 Placer	104	10	0	1	0	3	0	118
25 Merced	116	1	0	0	0	0	0	117

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

Stratum	Base Sample Landline	Cell Phone	Supplemental Samples					Total
				Surname*				
			Landline Geographic	Korean Only	Korean and Other	Vietnamese Only	Vietnamese/ Other	
26 Butte	72	16	0	0	0	0	0	88
27 Shasta	77	13	0	0	0	0	0	90
28 Yolo	110	12	0	0	1	2	0	125
29 El Dorado	77	2	0	0	0	0	0	79
30 Imperial	136	2	0	0	1	0	0	139
31 Napa	67	0	0	0	0	0	0	67
32 Kings	121	6	0	0	0	0	0	127
33 Madera	137	5	0	0	0	0	0	142
34 Monterey	106	24	0	1	0	0	0	131
35 Humboldt	70	4	42	0	0	0	0	116
36 Nevada	70	8	0	0	0	0	0	78
37 Mendocino	77	1	0	0	0	0	0	78
38 Sutter	88	13	0	0	0	0	0	101
39 Yuba	90	5	0	0	0	0	0	95
40 Lake	76	1	0	0	0	0	0	77
41 San Benito	130	4	0	0	0	0	0	134
42 Colusa, Glenn, Tehama	63	9	0	0	0	0	0	72
43 Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	51	11	0	0	0	0	0	62
44 Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	43	3	0	0	0	0	0	46

* Not drawn by sampling stratum.

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

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

				Supplemental Samples				
				Surname*				
Stratum	Base Sample Landline	Cell Phone	Landline Geographic	Korean Only	Korean and Other	Vietnamese Only	Vietnamese/ Other	Total
State	3,002	199	0	30	26	100	22	3,379
1 Los Angeles	604	18	0	12	10	15	0	659
2 San Diego	159	1	169	1	1	9	2	342
3 Orange	126	4	0	4	4	33	7	178
4 Santa Clara	88	4	0	2	3	25	6	128
5 San Bernardino	103	4	0	1	1	3	0	112
6 Riverside	121	7	0	0	0	1	1	130
7 Alameda	83	7	0	4	3	5	3	105
8 Sacramento	65	8	0	1	0	4	1	79
9 Contra Costa	51	7	0	2	2	1	0	63
10 Fresno	57	13	0	0	0	1	0	71
11 San Francisco	22	3	0	0	0	1	0	26
12 Ventura	70	8	0	0	0	0	0	78
13 San Mateo	31	3	0	1	1	0	0	36
14 Kern	51	1	0	0	0	0	0	52
15 San Joaquin	44	4	0	0	0	1	1	50
16 Sonoma	40	4	0	0	0	0	1	45
17 Stanislaus	38	4	0	0	0	0	0	42
18 Santa Barbara	43	3	0	0	0	0	0	46
19 Solano	40	8	0	1	0	0	0	49
20 Tulare	43	9	0	0	0	0	0	52
21 Santa Cruz	36	1	0	0	0	0	0	37
22 Marin	32	0	91	0	1	1	0	125
23 San Luis Obispo	25	1	0	0	0	0	0	26
24 Placer	32	5	0	0	0	0	0	37
25 Merced	49	2	0	0	0	0	0	51

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

Stratum	Base Sample Landline	Cell Phone	Supplemental Samples					Total
				Surname*				
			Landline Geographic	Korean Only	Korean and Other	Vietnamese Only	Vietnamese/ Other	
26 Butte	32	12	0	0	0	0	0	44
27 Shasta	25	4	0	0	0	0	0	29
28 Yolo	56	13	0	0	0	0	0	69
29 El Dorado	33	0	0	0	0	0	0	33
30 Imperial	51	1	0	0	0	0	0	52
31 Napa	22	1	0	0	0	0	0	23
32 Kings	41	1	0	0	0	0	0	42
33 Madera	43	1	0	0	0	0	0	44
34 Monterey	37	11	0	1	0	0	0	49
35 Humboldt	29	2	17	0	0	0	0	48
36 Nevada	27	4	0	0	0	0	0	31
37 Mendocino	35	2	0	0	0	0	0	37
38 Sutter	30	5	0	0	0	0	0	35
39 Yuba	40	4	0	0	0	0	0	44
40 Lake	41	0	0	0	0	0	0	41
41 San Benito	55	0	0	0	0	0	0	55
42 Colusa, Glenn, Tehama	32	3	0	0	0	0	0	35
43 Del Norte, Lassen, Modoc, Plumas, Sierra, Siskiyou, Trinity	19	5	0	0	0	0	0	24
44 Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne	24	1	0	0	0	0	0	25

* Not drawn by sampling stratum.

Source: UCLA Center for Health Policy Research, 2009 California Health Interview Survey.

Table A-6. Number of telephone numbers drawn in the cell phone sample by sampling stratum

Region	Stratum	Number of cell phone numbers drawn
State	State	58,900
1 - Northern & Sierra Counties	530	13,200
2 - Greater Bay Area	408	1,200
	415	1,200
	510	600
	650	600
	707	3,900
	925	1,200
3 - Sacramento Area	916	3,300
4 - San Joaquin Valley	209	3,500
	559	5,700
	661	500
5 - Central Coast	805	3,600
	831	3,800
6 - Los Angeles	213	1,600
	310	700
	323	1,800
	562	4,400
	626	600
	818	1,100
7 - Other Southern California	619	600
	714	600
	760	900
	858	900
	909	1,200
	949	1,400
	951	800