A Review of the Sample Design for the California Health Interview Survey

Ismael Flores Cervantes[1], Michael E. Jones[1], Laura Alvarez Rojas[1], J. Michael Brick[1], John Kurata[2], and David Grant[2]
Westat, 1650 Research Boulevard, Rockville, Maryland 20850[1]
UCLA, Center for Health Policy Research, 10960 Wilshire Blvd, Suite 1550, Los Angeles, CA 90024[2]

Abstract

This paper describes the sampling procedures used to select samples for the California Health Interview Survey (CHIS). CHIS is a telephone survey that has been conducted in 2001, 2003, and 2005. The objectives of CHIS are to examine issues in public health and health care and to monitor changes over time for Californians. Each round of data collection presented many challenges that had to be addressed at the sample design stage. Changes in data user needs and their effect on the sample design are discussed. Several sampling methods such as geographic stratification, oversampling of small areas, disproportionate stratified sampling for minority areas, and the use of surname list frames have been incorporated into the design. Techniques for increasing telephone dialing efficiency such as purging of nonresidential numbers, oversampling of cases with a matched address, and subsampling of cases for refusal conversion are also discussed.

Keywords: RDD, health survey, geographic stratification, surname list samples, within-household sampling

1. Introduction

The California Health Interview Survey (CHIS) is a telephone survey of California's population, conducted every other year since 2001. CHIS is the largest health survey ever conducted in any state and one of the largest health surveys in the nation. It is a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute. Funding for CHIS comes from multiple sources including federal government agencies and private foundations. Westat conducted the sampling and data collection in all three cycles of CHIS.

CHIS collects extensive information for all age groups on health status, health conditions and prevalence of chronic conditions, health-related behaviors, health insurance coverage, access to health care services, and detailed respondent demographics.

Data from each cycle of CHIS are available to a wide array of users such as state and local health agencies, community-based organizations, health care providers and organizations, advocacy groups, and policy makers through a variety of dissemination tools. The data enable researchers and policy makers to monitor the health of Californians and to examine changes over time.

Data from CHIS supports the production of estimates for the whole state, for the larger counties, and for groups of smaller counties in California. The survey also supports the study of the characteristics for the major, and a number of smaller ethnic groups in the state.

Each CHIS cycle required the use of multiple techniques to achieve statistically robust samples of the general population and of ethnic population groups. These challenges had to be addressed at the sample design stage. This paper describes the sample design for the first three cycles of CHIS. Section 2 compares the procedures used to design and select the sample of households across the collections. Section 3 presents the methods used for sampling persons within households. Section 4 discusses the response rates. The last section summarizes some of the issues and discusses the challenges for future cycles of CHIS.

Table 1 summarizes some of the key features of the sample design for CHIS by year. These features are discussed subsequently. More information on the survey design can be found in the Survey Methodology Report Series at www.CHIS.ucla.edu.

2. Sample Design

CHIS is a Random Digit Dial (RDD) telephone survey that selects and interviews one adult (age 18 years old or older) in each sampled household. In households with children (under age 12) or adolescents (ages 12-17) associated with the sampled adult, one child and one adolescent were sampled. A maximum of three interviews could be completed in each sampled household. The parent or legal guardian most knowledgeable about the health and care of the sampled child was interviewed. The sampled adolescent responded for him or herself, but only after a parent or legal guardian gave permission for the interview.

2.1 Mode Issues

A telephone sample was chosen for CHIS to meet the disparate needs of the survey. The large number of interviews and the state's geographic and linguistic diversity makes face-to-face interviewing cost prohibitive. The complexity of the data collection instrument necessitates using interviewers. Interviews were administered using Westat's computer-assisted telephone interviewing (CATI) system which controls the flow of the questionnaire depending on the answers provided and the information already known about the respondent.

Telephone surveys exclude persons in households with no landline telephones, including those in cell-only households and households with no telephone service. For estimates correlated with socioeconomic measures such as health insurance coverage, food security, and poverty, this coverage loss could introduce biases. In 2000, California had one of the highest rates of telephone coverage in the US with only 1.9 percent of households without a landline telephone while the proportion of households with only cellular service was very small. The proportion of non-landline households increased in the recent years due to more cell-only households. By 2005, about 10 percent of households were cell-only. Because households without a landline telephone were not sampled in the first three cycles of CHIS (other than a test in 2005), special weighting adjustments were implemented to reduce potential biases in the estimates. The increasing undercoverage presents challenges for future cycles.

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

Another source of coverage error in telephone surveys is the fact that persons who do not speak English are sampled but never interviewed because of language limitations[*]. In CHIS significant efforts have been made to limit bias from this source by interviewing in multiple languages including Spanish, Cantonese, Mandarin, Korean, and Vietnamese[†] (Lee et al., 2006).

2.2 Stratification and Sample Allocation

Two of the goals of CHIS 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 and (2) to produce reliable estimates for counties with populations of 100,000 or more and for groups of counties.

In CHIS 2001, the 58 California counties were grouped into 41 geographic sampling strata. Thirty-three of the 35 counties with a population of 100,000 or more corresponded to individual sampling strata. The two remaining counties with over 100,000 persons were each combined with an adjacent smaller county to form a stratum. The 23 remaining counties with populations of less than 100,000 were grouped geographically into the remaining six strata. The same stratification was used in 2003. The number of strata increased in 2005 to 44 in order to produce separate estimates for more counties. Eight previously combined counties were separated, leaving only three strata with more than one county (Table 1, row 1).

The need to produce reliable estimates at the county level required a compromise in allocating the sample. To achieve the most reliable statewide estimates, the optimal design is to allocate the sample to counties proportional to their population. For producing county-level estimates, the optimal design assigns each county an equal sample size. These allocations were not used because of the competing requirements. Instead, some of the sample in the largest counties under proportional allocation was re-distributed to smaller counties.

2.3 Race Ethnic Sampling

From the beginning, an important goal of CHIS has been to produce reliable estimates for as many race and ethnic groups in the state as possible (Table 1, row 2). The expected sample yield from the overall CHIS RDD sample was too small to support making inferences for many of these groups at the desired level of precision, so techniques were implemented to increase the yield to 500 completed adult interviews for some groups. In 2001 there were six such groups, but only the Korean and Vietnamese groups were targeted in 2003 and 2005.

---

[*] Language problem cases are considered nonrespondents in CHIS.

[†] Interviews in Khmer were conducted in CHIS 2001.

The race ethnic subgroups comprised a small percentage of the population, were geographically dispersed, and no single list contains all the members of the subgroup. As a result, sampling strategies for rare populations including disproportional stratified sampling and multiple frame sampling have been used to increase the sample of these groups (Kalton and Anderson, 1986).

Disproportionate stratified sampling was adopted in 2003 and 2005 for Korean and Vietnamese samples. Under this scheme, auxiliary information was used to classify telephone exchanges by the proportion of members of the target groups residing in these exchanges. After classifying the exchanges into strata, the telephone numbers in the exchanges with a relatively high proportion of members were sampled at a higher rate than the numbers in the other strata.

This geographic oversampling increased the sample yield for the race-ethnic groups in the RDD sample; however, the additional sample was not large enough to meet the goals for these groups. In order to meet the targets, samples from other frames (i.e., surname lists of the race-ethnic groups) have been used. In this approach, the RDD sample is supplemented with a much less expensive sample from a list of telephone numbers likely to include members of the target group(s). The list frame does not have to be complete to be useful, although the more complete the list the greater the potential for increasing the precision of the estimates. The composition of the list affects its efficiency (that is, the proportion of sampled numbers that lead to a member of the target group), but not the ability to produce unbiased estimates. Unbiased estimates can be produced if the list membership of every sampled unit (telephone number) from the other (RDD) frame 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 rare groups from the RDD sample. Although the use of surname lists was an effective way to increase the number of completed cases for these groups, the variances of the estimates for these groups is not greatly reduced by this approach.

The identification of "eligible" households (households with adults from the race or ethnic group of interest) in the list samples was done through screening. Screening eliminated unnecessary interviews because only eligible households were retained for further interviewing. If the household was ineligible, the screener interview was terminated. This strategy is relatively simple to implement and has good statistical properties, except for a measurement error that may be introduced when respondents misreport their race or ethnic group. Another disadvantage is that if the proportion of eligible households is small then a large number of households must be screened to achieve the goal. Screening was not necessary for the cases sampled from the strata with high/low concentration of Korean and Vietnamese since these cases are part of the RDD sample where all households are eligible for further interviewing.

2.4 Geographic Supplemental Samples

The second type of supplemental samples used is geographic samples added at the request of counties and local health departments interested in larger samples for a more detailed analysis. The geographic samples have covered different areas over the cycles (Table 1, row 3). The geographic supplemental samples that were whole counties were treated as regular RDD samples, but sometimes the samples were of smaller areas such as cities or areas within the cities. For supplemental samples of areas within a county, substrata within the county were created and sampled at different rates. Screening was not used in most geographic samples, but county-specific items were only administered for those that self-identified as being in the targeted county.

2.5 Special Supplemental Samples

The third type of supplemental samples in CHIS targeted special populations sometimes in specific areas. For example, in CHIS 2003, special supplemental samples were used to increase the number of African American adult interviews in the City of Hayward in Alameda County. In CHIS 2005 a state-level "child" supplemental sample was used to increase the number of child interviews in the state and in San Diego County (Table 1, row 2). As in the surname samples, screening questions were used to determine household eligibility. The same types of screening described above were applied.

2.6 Sampling Frames and Stratification

The frame of 100-banks with one or more listed telephone numbers was stratified into non-overlapping strata corresponding to a county or a group of counties. The geographic information required for stratification is available only at the exchange level. All banks within an exchange were stratified indirectly by mapping the exchanges to a county represented by the stratum. Exchanges that service households in more than one county were assigned to the county with the largest proportion of households serviced. Respondents might report living in a different county, but they were still interviewed. The misclassification between the sampling and self-reported strata affected the precision of the estimates.

Substrata were also created to draw samples for the supplemental geographic samples (i.e., a city or areas within a county), to control the sample size allocation within a county, and to oversample areas with high concentration of Korean and/or Vietnamese population in 2003 and 2005. Typically, the substrata were created using ZIP Codes of the areas serviced by the exchanges.

The assignment of telephone exchanges was sometimes problematic for small areas. If exchanges were assigned to the substrata that had the most telephone numbers, then they would cover a large proportion of households outside the target area leading to sample losses due to misclassification. On the other hand, if the substrata were created using exchanges that were contained within the cities, then the substrata covered only a proportion of the households in the cities leading to large design effects due to the different sampling rates.

Sometimes the creation of substrata could not be implemented prior to data collection because the funding for the supplemental samples was allocated after the first wave of telephone numbers were fielded. In this case, substrata were created after analyzing the distribution of the initial set of completed cases in the county. This required close monitoring of the sample yield of the initial sample to determine the exchanges with the highest proportions of the target population.

While disproportionate stratified sampling was used to oversample Koreans and Vietnamese, it did so without increasing the sample size allocated to any stratum. High and low concentration substrata were created in the four sampling strata where the Korean and/or Vietnamese population was large enough (in four counties covering about 78 percent of the Korean and Vietnamese population in California) to produce increases in the expected number of interviews. The high-density strata were exchanges with block groups with six percent or more Koreans or Vietnamese. The high-density substrata covered about 40 percent of the Korean and Vietnamese population and were subsampled at twice the rate of the low-density strata. This oversampling rate had only modest impacts on design effects.

After stratification, the sample size was allocated and the telephone numbers were selected. The number of telephone numbers to draw was adjusted to allow expected losses due to nonworking, nonresidential, and never answered numbers, nonresponse (screener and extended interviews), subsampling the numbers based on mailable status, and refusal status (in 2001 and 2003). These losses are typical of all RDD surveys except for the last two that are discussed below.

2.7 Increasing Data Collection Efficiency

Special data collection procedures are often implemented before data collection to reduce costs and to increase efficiency. One technique implemented in CHIS is the purging of unproductive numbers (i.e., business and nonworking numbers). Purging methods have been used in the all cycles of CHIS, with improvements in the methods over time. By 2005, about 50 percent of the sampled telephone numbers were purged (Table 1, row 5).

Another procedure implemented to improve the efficiency is a double sampling method. Telephone numbers are "nonmailable" (i.e., those without mailable addresses) are subsampled. Mailable telephone numbers are much more likely to be residential and to respond. The approach implemented the ideas presented in Brick, et al. (2002). Nonmailable telephone numbers were subsampled at a rate of 80 percent in 2001 and 75 percent in 2003. The subsampling procedure was not used in 2005 after re-evaluating the effectiveness of the method given the larger percentage of purged and mailable cases (Table 1 rows 5 and 6).

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

Screener interview refusals were subsampled at the rate of 60 percent in 2003 and 2005 (Table 1, row 8). This rate required a modest increase in the size of the sample of telephone numbers and causes less than a 3 percent increase in the standard error of the estimates. This technique was not used in 2001.

3. Within Household Subsampling

As mentioned before, one adult in the household was always sampled. A child and an adolescent were sampled if they were present in the households. Special within-household sampling procedures were developed

to maximize the analytic utility of the data collected from the respondents in CHIS.

One approach to sample persons within households is to simply list all the persons in the age group (adult, child, and adolescent) and select one person randomly from each group. This approach works well in most households since they have only one family. However, in households with two or more families, using this method could result in selecting persons from the different age groups who were not members of the same family. This situation is undesirable because the adult interview collected data about the family of only the sampled adult. The data from the adult interview are important for the analysis of the data from the child and adolescent interviews. Data from children or adolescents who were not members of the same family as the sampled adult have very limited utility.

Persons were sampled using a linked sampling approach where children and adolescents in the household were selected in two phases. In this approach children and adolescents in the household were linked to the sampled adult or his/her spouse or partner. That is, children and adolescents for whom an adult (or his/her spouse or partner) was a blood, adoptive, or foster parent or legal guardian were linked to or "associated" with that adult.

In the first phase of selection, an adult was randomly sampled from all the adults in the household. In the second phase, a child and/or adolescent was sampled from only the children/adolescents associated with the sampled adult. All other children/adolescents not associated with the sampled adult were ineligible for sampling. Since the sampling of children and adolescents was a two-phase procedure, the probability of sampling the child/adolescent was computed as the product of the probability of sampling the adult (phase one) and the probability of sampling the child/adolescent from all the children/adolescents associated with that adult (phase two).

One disadvantage of the linked sampling approach is that in a few households it was not possible to associate a child or adolescent to an adult because of unusual household structures. Consequently, any child or adolescent not associated with an adult did not have a chance of being selected. In CHIS 2001 unassociated children and adolescents were randomly associated to an adult in the household before sampling. This approach was abandoned in later cycles of the survey because explicit consent of the parent or legal guardian was required before they could be in the sample. Analysis of the CHIS 2001 data showed that the bias due to excluding unassociated children and adolescents was very small. Another disadvantage of the linked sampling approach was that no person in households with emancipated minors could be sampled because an adult was not present. These households were ineligible for sampling in all three cycles of CHIS.

Data needed for linking and sampling children and adolescents to the sampled adult and his/her spouse or partner were collected during the adult interview in 2001 and 2003 surveys.[‡] Beginning in 2005, children and adolescents could be sampled at the end of the screener interview when certain criteria were met (the screener respondent was the spouse/partner of the sampled adult, there were children in the household, and the sampled adult was not available at the time). When a child or adolescent was sampled at the end of the screener interview it was called the "child-first" procedure. The child-first procedure was an operational method (not a sampling method) used to increase the sample yield for children. In this case the child interview was conducted immediately with the screener respondent without waiting for the completion of the adult interview. Results of CHIS 2005 showed an increase of 16 percent of child interviews when compared to 2003.[§] The analysis also showed a difference of 15 percentage points in the overall response rates for child interviews conducted using the child-first procedure compared to those not using the procedure. Overall gains were limited because the procedure was used in only 30 percent of completed screeners with children.

3.1 Sampling Adults

The adult interview is the key interview in CHIS. Different sampling methods for selecting adults have been used in each cycle. In CHIS 2001, one adult per household was sampled using a modified Kish method that fully enumerated all adults in the household during the screener interview (Kish, 1949). Although in most cases adults were sampled with equal probability, some adults were selected with differential probabilities under special conditions. In households with adults both younger than 24 years old and 40 years old or older, older adults had twice the chance of being selected. This method reduced the chances of selecting adult children, thereby increasing the likelihood of selecting parents of children and adolescents in the household.

In CHIS 2003, a different method for sampling adults was implemented to reduce the intrusiveness of the

[‡] Data for linking the child/adolescent to the sampled adult's spouse were collected because the child could be selected though the spouse.

[§] Percentage adjusted to account the number of completed adult interviews in 2003 and 2005.

screening process. The method, the Minimally Intrusive method or Rizzo method (Rizzo, et al. 2004), avoids enumerating all adults in the vast majority of households. After asking the number of adults in the household, the next step is randomly deciding if the screener respondent is sampled. If the respondent is not sampled, then an adult is randomly sampled among the remaining adults in the household. In households with one or two adults, the method only requires asking the number of adults in the household. In households with three or more adults either the screener respondent is sampled or another adult is selected using the next-birthday method. Adults are selected with equal probability. A minor disadvantage of this method was that it requires sampling adults with equal probability, so the procedure could not be altered to reduce the likelihood of selecting adult children. It is difficult to measure the benefits of the Rizzo method without an experimental design (Table 1, row 10).

3.2 Child and Adolescent Sampling

Children and adolescents were sampled after the adult was selected. If there were any children under 12 in the household who were associated with the sampled adult, then exactly one child was sampled. Similarly, exactly one adolescent from all the adolescents associated with the sampled adult were sampled.

In 2001 and 2003, all eligible children and adolescents in the household were enumerated and sampled with equal probability of selection. During the enumeration it was determined if the children and adolescents were associated to both the sampled adult and his/her spouse or partner. Children or adolescents not associated with the sampled adult in a household were not eligible to be selected in this second phase of sampling. They were eligible to be selected if they were associated with other adults in the household. If the sampled adult did not have any associated child or adolescent, none was selected even if there was a child or adolescent in the household.

Since children and adolescents were sampled in two phases, the probability of selection for a child or adolescent was the probability of selection of the adult multiplied by the conditional probability that the child or adolescent was selected given that the associated adult was selected. If the child or adolescent was associated with two adults (sampled adult and his/her spouse/partner), the probability of selection was the sum of the probabilities calculated in this way for each adult. For example, the probability of sampling a child through the sampled adult was computed and added to the probability of sampling a child through the spouse/partner of the sampled adult.

As mentioned before, in CHIS 2005 a goal was to increase the sample of young children (0 to 5 years old). To increase the sample for younger children, a greater probability of selection was assigned for 0 to 5 year olds (Table 1, row 10). Comparing the results of CHIS 2003 and 2005, the proportion of completed child interviews age 0 to 5 among all child interviews increased from 46 percent to 51 percent. The number of cases for older children (age 6 to 11 years old) was reduced. The oversampling led to a slight increase in the design effect for estimates for all children.

4. Response Rates

Although response rates are not very useful for estimating nonresponse bias, they provide information on the success of the survey at representing the population sampled. The CHIS response rates are similar to those of other scientific telephone surveys in California, such as the California Behavioral Risk Factor Surveillance System survey (2005).

In general, it has become increasingly difficult for survey researchers to reach persons via landline telephones, and CHIS is no exception. In addition, survey response rates tend to be lower in California than nationally. Not only has contacting households become more difficult but there is an increasing reluctance of persons who have been reached to complete an interview.

A wide range of methods for increasing the response rates have been used in the cycles of CHIS. Some of these methods such as subsampling of refusals and the use child-first procedures have been described in the previous sections. Other methods implemented in CHIS were mailing advance letters (in five languages), monetary incentives, and different sponsorship for the advance mailing (Table 1, row 9). The effect of these methods is difficult to quantify because it is confounded with declining response rates (Table 1, row 13).

5. Summary and Future Challenges

In general, the sample designs, interviewing protocols, and operational procedures used during data collection in the three cycles of CHIS have been successful in achieving the goals of number of completed interviews. The procedures had to be modified to address the changing environment of telephone surveys, different objectives of the surveys, and budgetary constraints. Observed results such as the response rates from previous cycles have been used to refine the sample design of the following cycle. We expect the methods will need to continue to evolve in future cycles of

CHIS to ensure the quality of the data and its usability for researchers and policymakers in California.

Future cycles of CHIS and other RDD surveys face many challenges and obstacles. Two of the most important challenges are the declining response rates and the increasing noncoverage error due to the increased number of households with no landline telephone service. The use of methods described in the previous sections in combination with the development of innovative methods will be needed to address declining response rates. The increasing noncoverage error bias is the result of the increased popularity of cell phone use accompanied by a rise in the number of cell phone only households. Cell phones are not sampled in traditional telephone surveys. As the proportion of households with no landline increases, the bias associated with this group may be more substantial. Although past cycles of CHIS have not addressed this noncoverage error, a special area probability sample will also be selected in at least Los Angeles County in CHIS 2007. The area probability sample will provide the opportunity to better understand and address nonresponse and coverage in the next and future cycles of CHIS.

References

Centers for Disease Control and Prevention (2005). *Behavioral Risk Factor Surveillance System Survey Data*. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

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

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

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

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, 44, 380-87.

Lee, S., Kurata, J., Nguyen, H.A., and Jawad, M. (2006). *Shifting Focus from Nonresponse to Coverage Error: Linguistic Minorities in Surveys.* Paper presented at the annual meeting of the American Association for Public Opinion Research, Montréal, Canada.

Rizzo, L, Brick, J.M., and Park, I. (2004). A Minimally Intrusive Method for Sampling Persons in Random Digit Dialing Surveys. *Public Opinion Quarterly*, 68, 267-274.

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

Table 1. CHIS Summary

Reference Row	Description		Year 2001	2003	2005
1	Number of strata		33 single-county, 8 multiple-county	33 single-county, 8 multiple-county	41 single-county, 3 multiple-county
2	Race/ethnicity and special supplemental samples		State: South Asian, Cambodian, Japanese, Korean, Vietnamese. Urban/rural: American Indian/Alaska Natives Shasta Co.: Latinos	Oakland, Hayward, Alameda Co.: Korean, Vietnamese, African-Americans	Korean, Vietnamese, State-wide[1] child supplemental, San Diego child supplemental
3	Geographic supplemental samples		Counties: San Francisco, Santa Barbara, Solano Cities: Berkeley, Long Beach, Pasadena	LA-SPA: Antelope Valley County: Alameda Cities: Oakland, Hayward	LA-SPA: Antelope Valley Counties: Humboldt, Marin, Solano
4	Number of telephone numbers drawn		365,308	463,025	554,572
5	Purging method, percentage purged		Genesys ID, 24	Genesys ID Plus, 39	Genesys CSS, 45
6	Percentage with mailing address (after purging)		52	65	66
7	Nonmailable subsampling rate (%)		80	75	100
8	Refusal conv. subsampling rate (%)		100	60	60
9	Experiments		None	Address vendor	Letter head, Incentive
10	Within HH person selection method	Adult	Kish, unequal probability	Rizzo, equal probability	Rizzo, equal probability
		Child	Equal probability	Equal probability	Oversample 0-5 yr olds
		Adolescent	Equal probability	Equal probability	Equal probability
11	Average number of calls per completed interview[2]	Adult	3.8	3.7	3.7
		Child	2.3	2.5	2.6
		Adolescent	3.4	3.4	3.4
12	Actual number of completed interviews	Adult	54,122	42,044	43,020
		Child	12,392	8,526	11,358
		Adolescent	5,733	4,010	4,029
13	Response rates (%)	Screener	59.2	55.9	49.8
		Extended Household	63.7	59.9	59.3
		Adult	63.7	59.9	54.0
		Child	87.6	81.4	75.2
		Adolescent	63.5	57.3	48.5
		Adolescent[3]	84.5	83.3	77.5
		Overall Household	37.7	33.5	29.6
		Adult	37.7	33.5	26.9
		Child	33.0	27.3	37.4
		Adolescent	23.9	19.2	24.2
14	Coefficient of variation of the weights	Adult	96.6	98.3	114.9
		Child	105.5	102.2	128.1
		Adolescent	107.9	103.2	111.4

[1] Additional statewide and San Diego samples for child-first.
[2] Ignoring calls to nonrespondents etc.
[3] Cases where permission not granted removed from denominator.