SAMPLING RACE AND ETHNIC GROUPS IN RDD SURVEYS

W. Sherman Edwards, J. Michael Brick, and Ismael Flores-Cervantes, Westat;
Charles DiSogra and Wei Yen, UCLA Center for Health Policy Research

Introduction

Public policy researchers and decision-makers have long been aware that persons with different cultural backgrounds often have different needs with regard to health care, education, and social services. A number of barriers have prevented researchers from developing statistics, particularly survey-based statistics, for many cultural minorities, including the difficulty and expense of achieving sufficient sample sizes for separate subgroup estimates, and of preparing culturally appropriate data collection protocols. California, whose population recently became minority white non-Hispanic, includes among the largest and most diverse cultural sub-populations of any geographic political entity. This diversity presents challenges for state policy-makers and opportunities for policy researchers.

The 2001 California Health Interview Survey (CHIS) was designed to produce both geographic sub-state estimates and state-wide population subgroup estimates of health status and prevalence of chronic conditions, health-related behaviors, health insurance coverage, and access to health-related services. Because of the large number of interviews needed to support these goals, the complexity of the information being requested, and the need to keep data collection costs within reasonable bounds, researchers at the UCLA Center for Health Policy Studies and their partners selected random-digit-dial (RDD) as the sampling method for CHIS. The CHIS RDD sample design was expected to support precise local-area estimates for most California counties and for three cities with their own health departments, and state-level estimates for Latino, African-American, Chinese, and Filipino sub-populations. In addition to these groups, CHIS researchers were interested in producing separate estimates for Japanese, South Asian, Korean, Vietnamese, and Cambodian sub-populations, as well as urban and rural American Indians and Alaska Natives, and Latinos in Shasta County. Estimates for these groups required supplementing the expected RDD sample yield. In order to achieve the desired sample yields and enhance the representativeness of the survey for these subpopulations, the survey design also called for in-language interviewing in English, Spanish, Cantonese, Mandarin, Korean, Vietnamese, and Khmer.

Sampling Approach

The regular CHIS sample was selected using list-assisted RDD methods (Casady and Lepkowski, 1993) within each of 41 sampling strata. The sampling strata are either whole counties or combinations of smaller counties in California. 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 sampled by simple random or a systematic sampling. For CHIS, the sample was selected in two waves corresponding to an initial and final data collection period. In addition, telephone numbers that were listed or had a mailable address associated with them were sampled at a higher rate (about 1.25 times the rate used for the not listed numbers) using the methods described in Brick et al. (2002). Over 300,000 telephone numbers were sampled. Once the numbers were sampled, business numbers and some nonworking numbers were eliminated from the sample using automated procedures.

Both geographic and race/ethnic supplemental samples were selected for CHIS, but here we restrict our attention to the race/ethnic supplemental samples. As noted in the introduction, the race and ethnic subgroups are important for analytic reasons but constitute a small proportion of the total population and are dispersed throughout the state. As a result, the expected sample yield in even a survey as large as CHIS is too small to support making inferences for the subgroups at the desired level of precision. Because the members of these groups are a small percentage of the total population, are geographically dispersed, and no single list of all the members of the group is available, sampling methods for rare populations such as those described by Kalton and Anderson (1986) were considered for sampling these persons for CHIS.

Several sampling strategies were considered to increase the sample size for racial and ethnic subgroups in the CHIS 2001. The sampling strategies include household screening, use of auxiliary information to classify telephone numbers, network sampling, and the use of special lists. Screening works by sampling a large number of telephone numbers, retaining a household if it contains a member of the rarest subgroup, and subsampling otherwise. A second strategy uses auxiliary information to stratify telephone exchanges by the proportion of members of the groups residing in these exchanges and then samples the strata at differential rates. A third sampling strategy is multiplicity or network sampling, where each household sampled that belongs to the targeted group is asked to identify other households linked to them (by linkage rules such as siblings). The linked households are then interviewed. A fourth scheme is a dual frame design, in which the RDD sample is supplemented with a sample selected from a list of the telephone numbers for members of the groups.

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

In a dual frame approach, the list does not contain all the members of the group, but the characteristics of the list, including its completeness, are very important. Of course, the list must contain the telephone number for members of the subgroup. A second important property is the completeness of the list to make the procedure more efficient. A third property is the need to cover a relatively broad spectrum of types of members of the groups so the efficiency for different types of statistics for the subgroup can be improved. Finally, the list needs to be accurate in the sense of actually containing members of the group targeted; otherwise a large screening cost is incurred.

The lists for the supplemental samples were created using surnames of each of the race/ethnic groups. The only exception is the sampling for American Indians/Alaskan Natives discussed below. Genesys maintains a list of surnames associated with each of the groups targeted for CHIS and these were used to select a sample from the listed surname in the White Pages across the state. For the Shasta Latino sample, the list of Latino surnames was restricted to Shasta County.

The American Indian/Alaskan Native supplemental sample used a list was developed by UCLA in coordination with American Indian tribes and organizations. A large fraction of the listed telephone numbers was obtained from the U.S. Indian Health Services (IHS). The list was stratified into rural and urban strata and simple random samples were selected from each of the two strata.

Making the Design Operational

Screening for race/ethnic subpopulations. An important first step in making the design operational was defining membership in the subgroups of interest. The CHIS 2001 adult questionnaire included the standard Census questions on race and ethnicity, with added detail on Hispanic/Latino origin, on Asian and Pacific Islander subgroups, and on tribal membership for American Indian/Alaska Natives. For adults reproting more than one race or ethnicity, the inetrview also include a question asking for the group with which they most identified. Generally, individuals were eligible for the race/ethnic oversamples if (1) they reported only the target race/ethnicity or (2) most identified with the target race/ethnicity after reporting more than one. American Indian/Alaska Natives were also eligible if they reported being a member of a Federally recognized tribe, whether or not they most identified as American Indian/Alaska Native.

For the race/ethnic supplemental samples, one question was added after the enumeration of adults in the household, generally of the form:

> *Do any of these adults who live in your household consider themselves to be (ETHNICITY) or of (ETHNICITY) descent?*

If the answer to the ethnic screening question was "yes," then the interviewer asked whether each adult was of that ethnic background. Only adults of the appropriate ethnic background were eligible to be selected for the extended interview. In the extended interview, the sampled adult was asked about his/her racial and ethnic background. Those responding that they were something other than the supplemental sample category were also considered ineligible and the interview was terminated.

Two exceptions to these screening procedures arose during the field period. Fairly early on, the CHIS staff learned that Cambodians were very suspicious of telephone calls from strangers because of activities relating to political rivalries originating in Cambodia. The ethnic screening question was considered likely to put Cambodians on guard, and either refuse to respond to the survey or to answer untruthfully. The question was dropped for the Cambodian supplemental sample, and the screening took place in the extended interview. After the events of September 11, interviewers observed a similar phenomenon among the South Asian community. Soon thereafter, the ethnic screening question was dropped for the South Asian subsample.

In-language interviewing. Interviews were conducted in-language as needed for Latino, Vietnamese, Korean, and Cambodian (Khmer) respondents. The Spanish and Vietnamese translations were displayed in CATI, while the Korean and Khmer translations were on paper, with interviewers entering the responses in the English CATI version.

Bilingual interviewers were trained in English first, and then given interim assignments before interviewing in-language. Those with good English skills interviewed in English, while those with strong accents or limited English skills called RDD numbers identified as "language problems" to determine what language was spoken. Westat did not have Asian bilingual telephone center supervisory staff, so special procedures were developed for monitoring and supervision of the bilingual inetrviewers. During the interim assignments, a system of peer monitoring was developed under the direction of an experienced non-bilingual team leader, which was then implemented during in-language interviewing.

Sample yields. Table 1, displayed at the end of the paper, presents the sample yield for the RDD and each race/ethnic supplemental sample. The "sample yield" here means the ratio of completed eligible interviews to telephone numbers sampled. The overall yield is presented in the bottom row of Table 1; it is the product of the screener and adult interview yields in other shaded rows. Each of these yield rates is further decomposed to account for the sample loss due to known non-residency status (screener only), non-contact where eligibility was not determined, refusal to participate, and ineligibility.

The "Adult extended complete" numbers met or exceeded targets for all samples except the Cambodian list, where the yield rate was felt to be too low to warrant continuing, and the Shasta Latino list, where the list was exhausted. Highlights from Table 1 include:

- Generally, the non-residency rates were lower for the race/ethnic surname (all but AIAN) samples than for the RDD, but they were still in the 20-30 percent range despite being drawn from directories of listed numbers.
- The combined (screener and adult) eligibility rates varied considerably across the race/ethnic list samples, from a low of 24 percent for Cambodians to a high of 87 percent for Vietnamese.
- Cooperation rates were similar across race/ethnic list samples and the RDD;
- The combined yield rates for the race/ethnic samples were mostly lower than for the RDD because of the loss due to eligibility screening.

Relative cost per completed adult interview. Since there was no way to measure directly the cost per completed interview by sample, we modeled the level of effort required to complete data collection by sample by estimating the mean number of interviewer hours required to complete all of the instruments associated with one household. This estimate includes time spent interviewing, contacting respondents, and gaining cooperation for a particular case, as well as an amortization of time spent on nonresponse, ineligible, and out of scope cases. and of interviewer administrative time associated with project activities.The components of the model include the variation in interview length by language, the total number of calls made for cases in each sample, and the proportion of cases with child and/or adolescent interviews.

Spanish-language interviews took about 25 percent longer than the overall average, those conducted in Mandarin (RDD sample only) took about 50 percent longer, and those in Khmer about 75 percent longer. The other languages (Vietnamese, Korean, and Cantonese, RDD only) were in the range of 10 to 20 percent longer.

Despite these differences in interview length, the number of calls per completed adult interview dominated the model. The number of calls was a function of the sample yield as shown in Table 1, and also of the relative difficulty of contact. Variation in the proportion of cases with child and/or adolescent interviews had little effect on the results of the model.

The proportionate increase in level of effort compared with the RDD sample is shown in Table 2. The range is considerable – interviews for the Vietnamese sample required about the same level of effort as RDD interviews, while for the Korean sample it was more than twice as much and for the Cambodian sample more than 4 and a half times as much. The model does not account for other factors such as amortization of training and wage differentials for bilingual interviewers, so the level-of effort factors here probably understate the differences in total cost per case for the various samples.

Table 2. Proportionate increase in modeled level of effort, race/ethnic samples compared with RDD

Vietnamese	2%
Shasta Latino	18%
Amer Indian/Alaska Native	28%
Japanese	50%
South Asian	82%
Korean	123%
Cambodian	362%

Estimation Methods

The estimation and statistical inference schemes used for the regular RDD and geographic supplemental samples are standard, design-based approaches appropriate for large probability samples as described in survey sampling texts. These data and accompanying weights are appropriate for most CHIS analysis, including estimating characteristics of groups such as African Americans and Latinos that have relatively large sample sizes. Because of the special features of the race/ethnic supplemental samples, the data collected in the supplements the estimation weights for the target groups are identified separately from the data and weights of the regular RDD sample. Estimates for members of these specific groups will be produced from these data. The discussion below only discusses making estimates from the supplemental race/ethnic samples.

Two approaches to estimating characteristics of the race/ethnic groups were evaluated for each targeted group. One method is similar to that used for the regular RDD sample in that it assigns each household a base weight based on its probability of being included in the sample and makes standard nonresponse and population weighting adjustments to these base weights. We call this the *design-based* estimator in this paper. We refer to a second estimation approach as a *model-based* approach; with this method each household is assigned a constant base weight and these base weights are then adjusted. These methods are described in more detail below.

The design-based estimator is conceptualized by considering all telephone households in California as being in either on the supplemental list (L) or only as being eligible for sampling from the RDD sample (R). The base weight for a household that could only be sampled from the RDD is the ratio of N_R, the number of households in the state not on the list (actually in the sampling strata but we ignore this for the presentation), to the number of households sampled in the RDD that were not on the list (n_R). The base weight for households sampled that were on the list is the number of households on the list in the state (N_L) divided by the number of households on the list that were sampled in either the RDD or list (n_L). Duplicate telephone numbers were eliminated. Creating these weights required being able to classify every telephone number by whether or not it was on the list irrespective of how it was sampled. It is easy to show that the resulting weights are composite weights derived by averaging the RDD and list samples using a composite factor proportional to the sample sizes. Thus, the weight produces an unbiased estimate in the traditional design-based framework. The base weights are then adjusted for other steps of sampling such as within household sampling of adults and nonresponse. The final step is to poststratify the weights to the 2000 Census counts of the number of persons in the race/ethnic group, by age and sex categories.

The main problem with the design-based estimator is that the RDD and list base weights may be very different and this will result in estimates with very low precision. Kish (1992) describes the loss of precision due to differential weights. The model-based approach was developed to avoid the problem of having very different base weights by starting with the household base weights being equal for the RDD and list samples. The other adjustments used in the design-based method, including poststratification to Census control totals, are then implemented with these base weights. The model-based estimates are not unbiased in the traditional sense, but they should have lower variances than the design-based weights. The problem is that the bias of the model-based weights may be large if the persons on the list have characteristics that are different from the persons in the group that are not on the list. The evaluations below examine the potential bias by comparing characteristics of the sampled persons using the design-based and model-based weights. These differences are not reliable estimates of bias, but they do reveal whether the biases in the model-based estimates are likely to be very large.

The left-hand panel of Figure 1 graphs the estimates for a group computed using the model-based weights on the horizontal scale and the design-based weights on the vertical scale. If the estimates fall in close proximity to the main diagonal, then the bias of the model-based estimates is not large. The right-hand panel shows scatter diagrams of the corresponding standard errors estimated using jackknife replication methods. In this case, points near the main diagonal indicate that the design-based standard errors are not that much larger than the model-based ones and the two weighting approaches produce estimates with about the same effective sample size.

The first pair of graphs is for the Japanese sample(the Korean, Vietnamese, and AIAN samples look similar). While the estimates for the model-based and design-based procedures give estimates that are similar, the standard errors are also not very different for the two methods; the mean ratio of standard errors is 1.17. Whenever the standard errors for the design-based estimates are relatively close to those for the model-based estimates, we choose to use the design-based procedure to reduce the potential for bias in the estimates. For example, there are sizeable differences in educational attainment and marital status for the Japanese supplemental sample. To avoid concerns about biases due to the composition of the list, the design-based estimates are recommended.

We also recommend using the design-based procedure for the South Asian sample, shown in the middle set of graphs, even though the model-based procedure gives smaller standard errors as shown in the plot. The effective sample size for the model-based estimates is approximately 626 while the design-based effective sample size is 331. The South Asian sample is not as clear-cut as the previous samples, but the design-based procedure is preferred because of the potential for substantial biases in some model-based estimates.

The last set of graphs, for the sample of Latinos in Shasta county, present a different picture, with a model-based effective sample size of 223 and a design-based effective sample size of 44. (The Cambodian sample has similar, although not as extreme, values.) Even though the plots show some large differences between the model-based and design-based estimates, using the model-based weights may be beneficial because the sample size is very small with design-based methods. Users of the model-based weights should be aware of the possibility of large biases due to the composition of the list sample.

REFERENCES

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

Casady, R.J., and Lepkowski, J.M. (1993). Stratified telephone survey designs, *Survey Methodology*, 19, 103-113.

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

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

CHIS 2001 Methodological Reports will be available at *http://www.healthpolicy.ucla.edu/chis/l.*

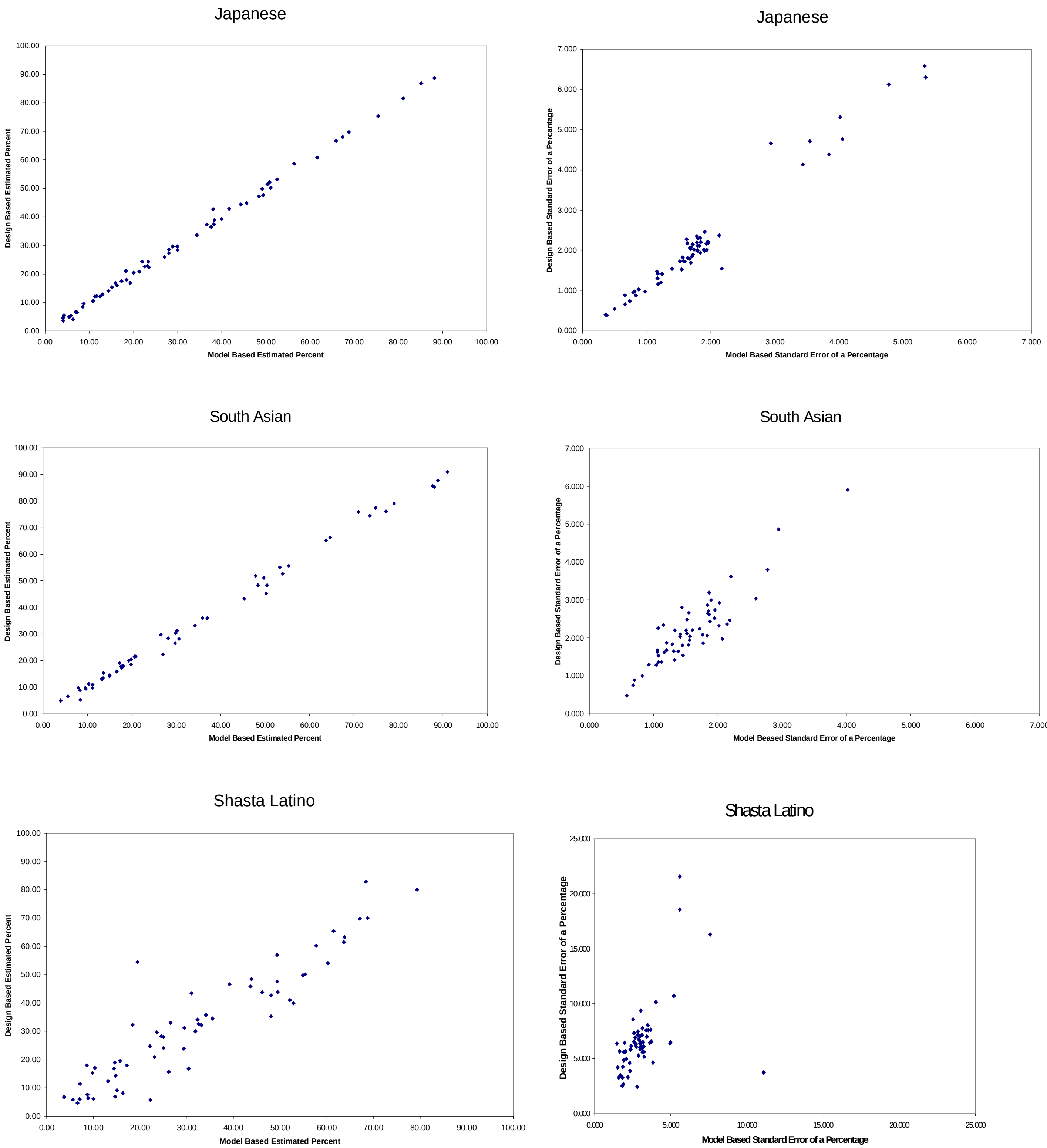

Figure 1. Scatter diagrams of model- and design-based estimates and standard errors for three supplemental samples.

Table 1. Sample Yield for CHIS RDD and Race/Ethnic Oversamples

	RDD[1]	AIAN	Cambodian	Japanese	Korean	South Asian	Vietnamese	Shasta Latino	All Samples
Initial sample size	309,111	2,953	2,567	2,465	3,639	3,673	2,984	1,906	326,345
Screener rates									
1 - Known non-residency rate	52.1%	57.2%	71.3%	78.0%	71.2%	70.9%	70.8%	69.3%	53.2%
Contact/resolution rate	73.4%	83.6%	68.1%	71.9%	68.2%	77.6%	72.0%	88.3%	73.4%
Cooperation rate	71.5%	76.7%	64.0%	64.2%	74.2%	65.6%	69.5%	77.9%	71.4%
Eligibility rate	99.3%	57.7%	73.3%	68.1%	39.3%	84.5%	91.9%	49.2%	97.1%
Screener yield	27.2%	21.2%	22.8%	24.5%	14.2%	30.5%	32.6%	23.5%	27.1%
Screener eligible	84,050	626	585	604	516	1,120	973	447	88,295
Adult extended rates									
Contact/resolution rate	83.9%	83.2%	79.0%	76.3%	76.4%	73.3%	74.3%	84.6%	83.5%
Cooperation rate	78.6%	80.8%	83.5%	75.9%	84.5%	81.0%	78.6%	82.0%	78.7%
Extended eligibility rate	99.9%	83.4%	32.6%	94.3%	97.9%	66.6%	95.1%	98.1%	99.0%
Adult extended yield	65.9%	56.1%	21.5%	54.6%	63.2%	39.6%	55.5%	68.0%	65.1%
Adult extended Complete	55,428	351	126	330	326	443	540	304	57,497
Combined rates									
Combined contact/resolution rate	61.6%	69.6%	53.7%	54.8%	52.1%	56.9%	53.5%	74.6%	61.3%
Combined cooperation rate	56.3%	62.0%	53.5%	48.8%	62.7%	53.1%	54.6%	63.9%	56.2%
Combined eligibility rate	99.3%	48.1%	23.9%	64.2%	38.5%	56.3%	87.3%	48.3%	96.2%
Combined yield	17.9%	11.9%	4.9%	13.4%	9.0%	12.1%	18.1%	15.9%	17.6%

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

[1]Includes county oversamples

Rate Definitions

1 - Known non-residency rate = (Initial sample - known non-residence) / (Initial sample)
Contact/resolution rate = (Completed eligible + Ineligible + Refusal) / (Total sample - Known non-residence)
Cooperation rate = (Completed eligible + Ineligible) / (Completed eligible + Ineligible + Refusal)
Eligiblity rate = (Completed eligible) / (Completed eligible + Ineligible)
Screener yield = (Completed eligible) / (Initial sample) = Residency rate x Contact/resolution rate x Cooperation rate x Eligibility rate
Adult extended yield = (Adult extended complete) / (Screener eligible) = Contact/resolution rate x Cooperation rate x Eligibility rate
Combined rate = Screener rate x Adult extended rate
Combined yield = (Adult extended complete) / (Initial sample)