

## **B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS**

### **B.1. Respondent Universe and Sampling Methods**

**Respondent Universe:** CHIS is a telephone survey of California's civilian, non-institutionalized population residing in households. The survey excludes potential respondents who are under 65 and too frail or ill to do the interview, unable to speak English or one of the four non-English languages in which the survey is offered, or hard of hearing. One adult (age 18 or older) and one adolescent (age 12-17) for whom the selected adult is the parent/guardian will be selected to participate in the survey. The target sample size is 48,000 adults and 4,000 adolescents. Additional information on the sample design is included in Attachment 7H, which shows the 44 geographic strata, the target sample size within stratum, the total number of households per stratum [based on California Department of Finance (CDOF) population projections for 2007], and the approximate unadjusted selection probability within each stratum.

**Sample Design and Sampling Methods:** The survey methods are consistent with the OMB Guidance on Agency Survey and Statistical Information Collections (January 20, 2006). CHIS 2007 uses a dual-frame sampling design. The first frame is a geographically stratified RDD sample of telephone numbers in California with a supplemental ethnic oversample. The second frame consists of cell-phone only households. The data from these two sampling frames will be integrated into a single data file in order to provide a more representative sample of California's non-institutionalized population.

The geographically stratified RDD sample is designed to produce both state-level estimates and county-level estimates for most of California's 58 counties. The sample is allocated to 44 geographic areas (sampling strata), defined as counties or aggregates of smaller counties with a minimum population size of 50,000 per stratum. A minimum sample of 500 is allocated to each stratum to maximize the effective sample size for county-level estimates and statewide estimates for major racial and ethnic groups. An ethnic oversample will supplement the RDD sample to provide robust estimates for Koreans (n=500), and Vietnamese (n=500).

The second frame of the CHIS sample will collect data on the small, but rapidly growing population segment of cell-phone only users. According to the February 2004 Current Population Survey, six percent of households had cell phones but no landlines. The purpose of the CHIS cell phone sample is to improve the coverage of the telephone survey and minimize any bias that could result from limiting the sample to residential households with landlines. CHIS 2007 proposes to include a sample of 800-1,000 cell-phone only households. A pilot study was conducted as an adjunct to CHIS 2005 to determine the feasibility of conducting CHIS with a sample of households with cellular telephone service only (no landline service). It showed that response rates for cell phone only users were similar to or slightly lower than those with landline phones, and that no practical limitations significantly influenced survey administration.

**Respondent Selection:** CHIS is a multi-stage interview -- first households are sampled and then respondents are selected from within households. At the screener stage, an adult informant (age 18 or older) residing in a household is contacted and asked how many adults reside in the household. If there is only one adult in the household, that adult is selected as the adult respondent. If there are two adults living in the household, the Computer Assisted Telephone Interview (CATI) software randomly selects one adult to be the CHIS respondent. If there are three or more adults, the screener informant will be asked which adult had the most recent birthday, and that adult will be the selected respondent. If the selected respondent is an older adult (65 years and older) who is too frail or ill to participate, the informant will be asked to identify a proxy for the selected older adult.

During the adult interview, adolescents age 12-17 residing in the household will be enumerated. Only if the selected adult respondent is the parent or guardian of one or more adolescents in the household will an adolescent be randomly selected. Following the completion of the adult interview, the interviewer will then attempt to contact the adolescent and ask him/her for assent to participate in the survey.

For the cell phone sample, the adult informant will first be asked whether he/she is speaking on a cell phone that is used solely for business purposes, and whether he/she has a landline in his/her residence. If the respondent answers "yes" to either of these questions, he/she

is ineligible for participation in the cell phone component. If there is only one adult in the household or there are multiple adults and each adult has a cell phone, then the adult answering the dialed number will be selected. If some members of the household share a common cell phone, then the CHIS sampling methods described above will be implemented to ensure that every adult in the household has an equal chance of selection. This randomization and selection approach yields a sample that is approximately representative of the adult and adolescent populations in each stratum in terms of characteristics such as age, gender, and race and ethnicity.

**Reporting Race/Ethnicity Data:** In all previous cycles of CHIS, OMB approved the method used to collect and report race/ethnicity data. First, CHIS collects race/ethnicity information in the question format mandated by OMB in the 1997 Revisions (OMB Bulletin No. 00-02, March 9, 2000). Respondents who report more than one racial group, or a racial group and Hispanic ethnicity, are subsequently asked if the respondent identifies "most" with a particular race/ethnicity. Consistent with previous cycles, the CHIS 2007 data set will include a race variable that is based on OMB standards for race/ethnicity and supplemental information about which race/ethnicity the multi-racial respondents most identify with, if any. Lastly, the data set will include a race/ethnicity variable created based on California Department of Finance standards. Because CHIS 2007 is co-funded by state, federal, and private funders, these additional race/ethnicity questions are needed to meet the requirements of its California sponsors. They do not conflict with either the collection of race/ethnicity information or the construction of variables based on the OMB standards.

**Response Rates:** In reporting response rates for CHIS, it must first be acknowledged that response rates for state-level surveys vary widely and are not comparable to response rates for national surveys. California as a whole, and the state's urban areas in particular, are among the most difficult in the nation in which to conduct telephone interviews.<sup>[3]</sup> In addition, California response rates have been decreasing, consistent with the national trend observed in other RDD surveys.<sup>[4],[5],[6]</sup>

Several dimensions of the survey methods used in CHIS make achieving high response rates particularly challenging. First, CHIS is an RDD telephone survey. A telephone survey is the only cost-effective mode for achieving the CHIS sample objectives of providing local level data and estimates for the state's major racial and ethnic groups. Similar surveys that are conducted in person, such as the NHIS, have higher response rates but produce relatively small samples and are far more costly. Second, as a population-based survey of households, virtually every household contacted by CHIS is eligible to participate in the survey. In other population-based surveys, only a small minority of contacted households is eligible to participate. Because the relative number of eligible households is much smaller (denominator) and the screening much simpler, they are able to obtain higher response rates.

Comparing survey response rates is further complicated by the use of different methods of calculation. Both the Council of American Survey Research Organizations (CASRO) and the American Association for Public Opinion Research (AAPOR) have developed standard methods for calculating response rates; however, there is considerable variation in how these formulas are implemented. The central problem is the difficulty in resolving the eligibility of the sampled telephone numbers that are never answered. Differences in disposition codes used by various CATI programs, different methods for allocating responses to eligibility categories, and different cut-off points for coding an interview as complete all contribute to variation in response rates. In addition, some surveys report weighted response rates and others unweighted rates. And multi-stage surveys differ in how they incorporate the screener and extended interviews into the response rate formula.

The CHIS 2005 response rates are comparable to those of other scientific surveys in California. The CHIS 2005 response rates have been calculated consistent with the standard approach used by the CDC BRFSS. The CHIS 2005 disposition codes were matched with those reported in the *2005 BRFSS Summary Data Quality Report*. For CHIS 2005, the Screening Response Rate, the proportion of all known households in which the presence or absence of an eligible respondent has been determined and in which an interviewer *actually spoke* to the selected respondent, was 54 percent, compared with 49 percent for the BRFSS in California. The

CHIS 2005 Extended Interview Response Rate, the proportion of contacted selected respondents who successfully completed an interview, was 63 percent as compared with 67 percent for the California BRFSS. The cooperation rate was 62 percent, compared with 59 percent for BRFSS. Because BRFSS reports the screener and extended interview response rates as a single unit (rather than multiplying the screener by the extended interview rates to calculate overall response rates), it is not possible to calculate comparable overall rates for CHIS.

Moreover, a survey's response rate is not the only, or even the best measure of its quality. One proven way to assess a survey's representativeness is to compare its findings with those of other similar surveys. An experiment conducted by the Pew Research Center in 2003 compared two surveys that used differential levels of interviewing effort on 90 measures. The surveys had response rates of 51 percent and 27 percent respectively.<sup>[7]</sup> The results found little difference between the two polls. Other studies of polls and omnibus surveys support the finding that low response rates do not cause nonresponse bias.<sup>[8]</sup> To assess CHIS validity, CHIS 2003 data were benchmarked to estimates on key health insurance, health care access, and health status indicators from the NHIS California sample. The benchmarking study was undertaken in collaboration with the National Center for Health Statistics (NCHS). The CHIS data are collected through telephone self-reports; however, the NHIS data are collected in-person and have a higher response rate, providing a relative "gold standard" to compare with CHIS estimates. Initial analysis of CHIS and NHIS data found the estimates of demographic and socio-economic variables were comparable. Although the benchmarking study did find some differences in specific health indicators, there were no indications of systematic bias. A similar comparison of CHIS and BRFSS key health estimates showed consistent results.

In CHIS 2007, a number of proven strategies to maximize the response rates will be implemented; these efforts are documented in B.3. Furthermore, to estimate survey bias due to non-response and undercoverage of households without telephones, an area probability pilot study with in-person recruitment is proposed for implementation in Los Angeles County. Additionally, the CHIS 2005 benchmarking study will be refined and expanded.

---

[1] Tucker, Clyde, J. Michael Brick, and Brian J. Meekins. 2005. "Telephone Service and Usage Patterns in the U.S. in 2004: Implications for Telephone Samples." Paper presented to the Committee on National Statistics, Washington, DC.

[2] 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.

[3] Behavioral Risk Factor Surveillance System, 2005. Summary Data Quality Report. California ranked 43rd out of 52 state reporting units in overall response rates and 48th out of 52 in the Council of American Survey Research Organizations (CASRO) response rates for the 2004 Behavioral Risk Factor Surveillance System.

[4] Curtin, R., S. Presser, and E. Singer. 2003. Recent Response Rate Changes in the Michigan Survey of Consumer Attitudes. Paper presented at 2003 meetings of the American Association for Public Opinion,

[5] Keeter, S., J. Best, M. Dimock, and P. Craighill. 2004. The Pew Research Center Study of Survey, Nonresponse: Implications for Practice. Paper presented at 2004 meetings of the American Association for Public Opinion.

[6] Curtin, R., S. Presser, and E. Singer. 2005. Changes in Telephone Survey Nonresponse Over the Past Quarter Century. *Public Opinion Quarterly*. 69(1), 87-98.

[7] The Pew Research Center for the People and the Press. 2004. Polls Face Growing Resistance, But Still Representative: Survey Experiment Shows.

<http://people-press.org/reports/display.php3?ReportID=211>

[8] Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64:413–28; Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a Large National Telephone Survey." *Public Opinion Quarterly* 64:125–48; Merkle, Daniel, and Murray Edelman. 2002. "Nonresponse in Exit Polls: A Comprehensive Analysis." In *Survey Nonresponse*, ed. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, pp. 243–58. New York: Wiley.

## **B.2. Information Collection Procedures/Limitations of the Study**

**Survey Introduction:** CHIS data will be collected via telephone interviews from civilian, residential households in California. The RDD sample frame will be matched against list directories, using reverse directory services, to obtain address information so that an advance letter can be mailed to potential respondent households to explain the purpose of this study (see Attachment 4). The advance letter will be mailed to all non-cell phone respondents, about 75 percent of the households in the CHIS sample. The CHIS 2007 sample of 800-1000 cell-phone telephone numbers cannot be matched to addresses; therefore, cell phone households will not receive advance letters.

**Survey Administration:** CHIS 2007 interviews will be administered as an RDD survey through a CATI system by interviewers trained by the data collection contractor and CHIS staff. CHIS data will be collected over a 6 - 9 month period to distribute the data collection burden and to minimize any seasonal biases.

CHIS interviewers will receive at least 18 hours of project-specific instruction in addition to the general interviewer skill training and CATI skill training provided to new interviewers. In addition, each interviewer will receive four hours of refusal avoidance training that focuses on providing answers to frequently asked questions, voice quality, and listening skills. Periodically, interviewers will also receive refresher training.

To minimize data entry errors, data consistency checks and range checks will be built into the CATI programming for CHIS. To ensure quality in the interviewing process, interviews will be randomly monitored via telephone from a remote station throughout the data collection period. All CHIS telephone calls made by the interviewers will be logged daily in detailed tracking reports, which will routinely be reviewed for irregularities and as a check on progress. In addition, a behavioral coding project will be implemented during the 2007 CHIS to evaluate the performance of questions on the adult questionnaire that measure racial-ethnic discrimination. Trained coders will listen directly to a sample of approximately 1440 interviews administered in all survey languages.

### **B.2.1. Statistical Methodology for Stratification and Sample Selection**

**RDD Sample:** CHIS uses an RDD telephone number generation technique that uses 100-banks with one or more listed telephone numbers to create a sample of potential residential households within each stratum. This produces a selection probability for a household that is equal to the ratio of the number of households selected into the sample over the total number of households known to exist in a stratum. Additional information on the sample design is included in Attachment 7H, which shows the 44 geographic strata, the target sample size within stratum, the total number of households per stratum [based on California Department of Finance (CDOF) population projections for 2007], and the approximate unadjusted selection probability within each stratum. To create the Korean and Vietnamese oversamples, CHIS employs geographic oversampling in areas of high concentration of these subgroups and also samples from a surname list sample. The interviewer confirms the ethnicity of each respondent whose telephone number comes from the surname list sample prior to enrolling the respondent in the survey.

**Cell Phone Sample:** The cell phone sample will be drawn from a statewide RDD sample of cell phone numbers from 1000-blocks in California that are cellular (NXXTYPE types 04, 55, 60) or PCS (types 65, 68)[\[2\]](#).

### **B.2.2. Estimation Procedure**

CHIS 2007 data will be statistically weighted to account for the differential probability of selecting persons into the sample, and the weights will be raked to the various domains of California population totals. The ethnic surname list sample and cell-only sample will be combined with the RDD sample and weighted together.

The estimation procedure will first weight the data on the probability of household selection. Adjustments will be made for households without telephones. Then, the weights of households with more than one voice line will be adjusted to correct for their greater than normal probability of selection. Next, the person-level weight will be created by multiplying the adjusted household weight by the number of adults in a household. A post-stratification estimation

procedure will then be performed to the person-level weight to bring the sum of weights to the total adult population using CDOF data projections for CHIS 2007 that are based on the 2000 Census data. Seven variables will be used in the post-stratification procedure to determine the final person weight: age, gender, race, ethnicity, geographic stratum (i.e. city, county, strata, and state), education, and home ownership.

The ethnic supplemental sample will be combined with the RDD sample and weighted together, using the dual-frame method developed for CHIS 2003 where the base weight accounts for the multiple selection probabilities for samples drawn from both the RDD and the surname list. The weighting procedure for the cell-only sample will require modeling because there is no reliable data source that provides totals and characteristics of the cell-only population in California. NHIS data will be used to build the model and the CDOF data will provide the control totals. This will provide a CHIS data set that represents the California's population with the smallest undercoverage and nonresponse error possible for the proposed design.

### **B.2.3. Degree of Accuracy Needed for the Purpose Described in the Justification**

CHIS is used for estimates of disease prevalence, program participation, health behaviors, insurance status, etc., for individual counties, race/ethnic groups and other subpopulations of interest (e.g. the elderly) in the California population. The large sample size allows robust estimates for any subpopulation with a sample size of 450 or more with a margin of error of less than .5 percent. For gender, race, ethnicity, or age, estimates at the state level can be obtained with a margin of error of less than 5 percent. At the county/stratum level, the minimum sample size of 500 will produce estimates with a margin of error at or below 7 percent, even with split male/female analyses. In short, CHIS estimates should approximate the California population.

### **B.2.4. Unusual Problems Requiring Specialized Sampling Procedures**

CHIS 2007 implements specialized sampling techniques for the cell phone sample frame and the area probability sample frame in Los Angeles County. The cell phone sampling techniques are discussed more extensively in B.1., and the area probability sample in B.3.

#### **B.2.5. Use of Periodic (Less Frequent Than Annual) Data Collection Cycles**

CHIS 2007 CCM is proposed as a one-time data collection.

#### **B.3. Methods for Maximizing the Response Rate and Addressing Issues of Nonresponse**

In CHIS 2001, CHIS 2003, and CHIS 2005, a number of generally accepted techniques were used to maximize response rates. As an initial strategy, CHIS uses an advance letter to differentiate CHIS from telemarketers. The advance letter (Attachment 4) explains the purpose of the survey, the sponsors, and its importance, as well as assuring potential respondents that their participation in the survey is voluntary and that their confidentiality will be protected. In CHIS 2005, 66 percent of households were mailed an advance letter and these households had a screener response rate almost 12 percentage points higher than the "no-letter" households. Because having an address is highly related to screener response rates, the data collection contractor is working to further improve its ability to match telephone numbers with addresses.

Other techniques used to increase response rates in CHIS 2001, CHIS 2003, and CHIS 2005 will be repeated in CHIS 2007, including: leaving a message on answering machines (only on first encounter) to announce the survey; dialing a non-responding telephone number at least 14 times over a range of time periods (daytimes, evenings, weekends, etc.); and providing a toll-free number for respondents to call back and set an interview appointment time.

Mailing a "refusal conversion" letter to households that do not firmly decline an initial invitation to participate has also been effectively employed in national RDD surveys as a way to convert these households to participate in the survey. In CHIS 2001, CHIS 2003, and CHIS 2005, this method helped convert about one-third of reluctant households. The method will be

implemented again in CHIS 2007. If a mailing address is available, a letter will be mailed to the household asking them to reconsider and restating the importance, legitimacy and purpose of the survey. The potential participant will then be re-contacted to provide an additional opportunity to participate in the study. Specially trained interviewers will make refusal conversion telephone calls. Sample refusal conversion letters are in Attachment 4.

To maximize participation among California's diverse ethnic populations, the CHIS 2007 will be administered in five languages: English, Spanish, Chinese, Korean, and Vietnamese. Building on materials previously translated for the CHIS 2001, CHIS 2003, and CHIS 2005 questionnaires, new questions are translated and reviewed for cultural adaptation. Specially trained bilingual/bicultural interviewers will conduct non-English interviews.

To increase interviewer's skills at encouraging individuals to participate in the survey, CHIS 2007 training, coaching, and monitoring will be intensified. The CHIS 2007 training will focus on introducing the survey and handling reluctant or difficult to reach respondents.

CHIS 2005 implemented a pre-paid \$2 financial incentive, which increased response rates by three percent. CHIS 2007 will also include pre-paid financial incentives of \$2.00 in the advance letter sent to all households with an available address. This result is consistent with other research, which indicates that pre-paid incentives result in more interviews, more appointments, and lower resistance. [\[1\]](#),[\[2\]](#)

By implementing these approaches, we expect to achieve an approximate 70 percent Extended Interview Response Rate and a 60 percent Screener Response Rate for CHIS 2007. However, a survey's response rate is not the only, or even the best measure of how well the survey estimates represent the target population.

**CHIS Benchmarking Study:** To evaluate potential survey bias, CHIS will build upon the CHIS 2005 benchmarking study (see B.1.) by using modeling techniques to compare CHIS and NHIS data estimates. In collaboration with the NCHS, UCLA has prepared a data set combining 2003 CHIS data and 2003 and 2004 NHIS data for California. The combined CHIS /NHIS data will be used to draw substantive conclusions about potential survey biases. The study will use a

multivariate modeling approach that examines the effect of one variable at a time, while controlling for other variables, on bias in CHIS health estimates.

**CHIS In-Person Area Probability Pilot Study:** To further evaluate the nature and magnitude of bias due to nonresponse to the telephone survey and undercoverage of households without telephones, CHIS 2007 will field an area probability sample that uses in-person recruiters. The sample will be fielded in Los Angeles County, California's largest urban area with the lowest response rate. The goal of the pilot test is to evaluate whether or not CHIS estimates are significantly biased due to nonresponse and RDD sample frame undercoverage. If significant bias is detected, strategies to alleviate the bias will be explored for future CHIS data collection cycles.

The initial sample will be drawn from a list of residential addresses in Los Angeles County, based on U.S. Postal Service Delivery Sequence files. A CHIS vendor will then attempt to match each address with a telephone number, using reverse directory services. To produce the final sample, addresses with telephone numbers will be subsampled at twice the rate as addresses without telephone numbers. The households without telephone numbers will be contacted in person. The households with telephone numbers will first be contacted by telephone, using the standard CHIS interview protocol, and, if necessary, will be followed up in person.

In person household contact will be attempted for the following types of cases:

- Sampled addresses without matched telephone numbers;
- Telephone refusal cases (only "soft" or nonhostile refusals, either to the screener or extended interview); and
- Maximum call cases (ring no-answer cases or cases where inconclusive contact has been made with a household member and the maximum allowable call limit was reached).

The target sample size for the area probability sample will be 800 – 1,000 adults in Los Angeles County. About half of these interviews are expected to be completed through in-person recruitment. To limit introducing mode differences, the in-person recruiters will direct a CHIS telephone interviewer to initiate a call to the selected respondent's landline telephone (if available) or alternatively to a cell phone that the recruiter will provide for use by respondents in households without landline telephones.

CHIS will offer \$25 to respondents to encourage participation in the study when in-person contact is made. This incentive amount is comparable to those of national surveys with federal sponsorship, such as the National Household Education Survey (NHES: 2007, OMB No. 1850-0768, Federal Register: September 18, 2006, Volume 71, Number 180, pp. 54628-54629) and the National Survey on Drug Use and Health (NSDUH, OMB No. 0930-0110, Federal Register, June 8, 2006, Volume 71, Number 110, pp. 33311-33312).

The weighting for the area probability sample will follow the same general approach used for the RDD and supplemental samples. The base household weight will be calculated as the inverse of the probability of selecting the address from the address sampling frame. Lastly, the samples from these different sampling frames will be weighted to the California population totals.

A minimum of 7,000 RDD and 800 area-probability sample respondents will provide the capability of detecting bias as small as 5 percent at a  $p\text{-value} \leq 0.20$  in most CHIS estimates.

[1] Brick, J. M., Hagedorn, M. C., Montaquila, J., Roth, S. B., and C. Chapman. 2003. Monetary Incentives and Mailing Procedures in a Federally Sponsored Telephone Survey. U.S. Department of Education, National Center for Education Statistics.

[2] Cantor, D., Cunningham, P., Triplett, T., and R. Steinbach. 2003. Comparing Incentives at Initial and Refusal Conversion Stages on a Screening Interview for a Random Digit Dial Survey.

#### **B.4. Tests of Procedures or Methods**

Most CHIS 2007 CCM questions are adopted from previous NHIS Cancer Supplements. A few original questions were constructed and cognitively pretested. For these reasons, questions used in the CHIS 2007 CCM are expected to produce reliable data.

The English version of the final draft instrument was pre-tested with nine persons (the OMB maximum prior to approval). Due to the small number of subjects, the pretest was conducted as an interviewer administered, telephone interview with a paper and pencil instrument rather than a CATI system. These pretests check the flow, clarity, difficulty level, and cultural bias of the questions.

The instrument also will be submitted to a CATI pilot test before it is fielded. The pilot test will test the adaptation of the instrument to the CATI system. A total of 150 pilot test interviews are currently planned after OMB approval is obtained. After the first round of pilot testing, the final English version will be translated into other languages and subsequently pilot tested in each language in which CHIS is offered.

## **B.5. Names and Telephone Numbers of Individuals Consulted**

As described in A.8, a Sample Design and Survey Methods TAC, consisting of statisticians and survey experts, provides expert advice to CHIS on the weighting schemes, imputation methods, and analytical plans. Members of this TAC are listed in Attachment 7D. In addition, a survey mode planning workgroup that included national experts convened to propose survey design options for measuring survey bias in preparation for CHIS 2007 (see Attachment 7F). The survey data collection subcontractor for CHIS 2007, Westat Inc., was chosen through a competitive bidding process at UCLA. Westat, Inc. has extensive expertise in survey methodology and has conducted numerous major federal surveys. As described in A.2. (Purpose and Uses of Information), CHIS data is widely used by state and federal agencies, county health departments, universities, research organizations, advocacy groups, community organizations, health care providers, doctoral students, and others. Attachment 7H provides lists of organizations that have used CHIS data and peer-reviewed peer publications based on CHIS data, as well as descriptions of the types of research conducted.