
THE 2003 CALIFORNIA HEALTH INTERVIEW SURVEY SAMPLE WEIGHTS

INTRODUCTION

The California Health Interview Survey (CHIS) data files—and many other survey data files with complex sampling designs—include weights to allow researchers to create population estimates. Weights are variables added to the data files that can be applied using statistical software programs, such as SAS, STATA and SPSS. As their name implies, these variables *weight* the survey data to approximate the California population living in households.¹ There are three samples in CHIS, and data files containing separate weights have been created for each sample.

1. Adults—persons age 18 years and older living in households in the state of California. The Adult weight variable in the CHIS 2003 data file is named **RAKEDW0**.
2. Children—persons age 0 through 11 living in a household with their mother and/or father or legal guardian. The Child weight variable is **RAKEDW0**.
3. Adolescents—persons age 12 through 17 living in a household with their mother and/or father or legal guardian. The Adolescent weight variable is **RAKEDW0**.

When the weight variable in the CHIS 2003 Adult survey data file is applied, for example, the data estimates approximate the population of all adults in California living in households as projected by the California Department of Finance in 2003 (25,597,062).

This brief document provides some general information on why you may want to use the weights, provides an example of estimates produced with unweighted and weighted data, describes how the weight variables were created, and finally describes the “*replicate weights*” that are also included with the CHIS 2003 data files.

¹ Adults, children and adolescents may live in either Households or Group Quarters (institutionalized and non-institutionalized). Persons living in Group Quarters, as defined by the Census Bureau, are not eligible to participate in CHIS. Common examples of Group Quarters include prisons, nursing homes, hospitals and dormitories.

WHY USE WEIGHTS?

To produce estimates from the CHIS 2003 data files, each completed adult, adolescent and child interview is assigned a sampling or estimation weight. The weights must be used with the data from the interviews to produce population estimates from the survey. The weights are needed because of the complex nature of the CHIS 2003 sampling and estimation methods. If the weights are not used, then the estimates will be biased and not give reliable estimates of the California populations.

One way of thinking about the sampling weight assigned to a record is that it is the number of individuals in the population that are *represented* by the respondent's answers to the survey questions. The weights are constructed so that the sample represents the California household population by county and by a number of demographic characteristics. Weighting compensates for factors in the survey design and implementation that would otherwise skew or bias estimates made from the survey data.

AN EXAMPLE OF UNWEIGHTED AND WEIGHTED ESTIMATES

As an example of how the weights are used with CHIS 2003 data, consider reports of a cancer diagnosis. The risk of cancer increases with age. For a variety of reasons, older persons are more likely to complete an interview than are younger adults. This means that older people are overrepresented in the unweighted CHIS data, and health conditions related to age will likewise be overrepresented. Table 1 shows CHIS 2003 responses to the question, "Has a doctor ever told you that you had a cancer of any kind?" The unweighted column shows that 4,727 adult respondents (11.2%) said "Yes" to this question out of all adults interviewed (42,044). This unweighted estimate indicates that 11.2% of the CHIS sample (*not* the population, which is all adults in California households) reported a doctor's diagnosis of cancer. The weighted column in Table 1 shows the population estimate that is produced when the weights are applied to the sample data. The weights adjust for the varying probabilities that people of different ages, gender and race/ethnic groups will be selected, and also adjust for the overrepresentation in the sample of older adults (bias). The weighted sample data yields the estimate that the proportion of the population of adults in California living in households (25,597,062) who report a cancer diagnosis is 8.3% (or 2,127,075), nearly three percentage points lower than the unweighted proportion (11.2%).

Table 1: Has a Doctor Ever Told You That You Had a Cancer of Any Kind?

Response	Unweighted		Weighted	
	Number	Percent	Number	Percent
Yes	4,727	11.2	2,127,075	8.3
No	37,317	88.8	23,469,986	91.7
TOTAL	42,044	100.0	25,597,062	100.0

Source: 2003 California Health Interview Survey

HOW THE CHIS WEIGHTS ARE CONSTRUCTED

Construction of the CHIS 2003 sampling weights is a complex and iterative process. A brief summary of the main steps follows. For the complete details on the CHIS weighting process, please see *Methods Report 5, Weighting and Variance Estimation*, available on the CHIS Web site at <http://www.chis.ucla.edu/methods.html>.

CHIS 2003 begins with a sample of telephone numbers selected for each of 41 counties or groups of small counties; these geographic divisions are called sampling *strata*. The total CHIS sample was allocated to these strata to support separate estimates for each stratum. As a result of this allocation, the probability that a given telephone number would be selected varied considerably by stratum. The first step in the weighting process is to create a *base weight* for each telephone number. The base weight is calculated by dividing the number of telephone numbers in a stratum by the total number of telephone numbers sampled from that stratum. The base weight is then adjusted in a series of steps to produce the final weights.

The first series of adjustments is conducted at the *household* level, and is associated with the screening portion of the CHIS interview. The screening process determines whether the telephone number is for a residence and whether there are other telephone numbers in that household, identifies a sampled adult, and determines whether the household includes children and/or adolescents. Only telephone numbers for sampled households that are residential and that respond to the screening process are assigned weights at this stage.

Other adjustments at the household level compensate for other features of the sample design, such as whether an advance letter was sent to the household and whether an attempt was



CALIFORNIA HEALTH INTERVIEW SURVEY
UCLA CENTER FOR HEALTH POLICY RESEARCH
10911 WEYBURN AVENUE, SUITE 300
LOS ANGELES, CA 90024
CALL TOLL-FREE: 1-866-275-CHIS (1-866-275-2447)
FAX: (310) 794-2686
WEBSITE: www.chis.ucla.edu

~~made to gain cooperation among households who initially declined to participate. A third design feature of CHIS 2003 is that part of the sample was screened for people of Vietnamese or Korean~~

background in order to increase the sample size for these groups. In these parts of the sample, households not including an adult of the target ethnic group were dropped and the weighting process adjusted for these households as well. A final adjustment to the household-level weight compensates for the fact that survey response rates are different for households with children and those without children. A post-stratification adjustment was used to account for this differential.

The final household weight is the starting point for deriving the adult, adolescent and child weights. The process is very similar for each of these samples. For the adult weight, the first step is to multiply the household weight by the number of adults in the household, since only one adult is selected from each household. The adult, adolescent and child weights were then subjected to one more weighting adjustment to make CHIS 2003 estimates consistent with the 2003 California Department of Finance (DOF) Population Projections. *Raking* is an estimation procedure similar to post-stratification, but it controls to several different known population totals. Basically, the weights are post-stratified to one set of population control totals (a dimension), and then these adjusted weights are post-stratified to another dimension. The procedure continues until all dimensions are adjusted. The process is then iterated until the control totals for all dimensions are simultaneously satisfied. The dimensions were created from variables common to the CHIS 2003 and DOF projections, including age, gender, stratum and other geographic levels, race and ethnicity. A set of dimensions was also included to adjust for households that do not have telephones or have only cellular telephones, since these households could not be reached through the sample of landline telephone numbers. Variables used in the dimensions were housing tenure (own or rent), categories of age and education of the sampled adult, and household size.

REPLICATE WEIGHTS

The sampling and estimation techniques used in CHIS 2003 are complex and, as a result, the standard errors of the estimates should not be computed using software that assumes the data are from a simple random sample. Computations assuming a simple random sample will generally underestimate the standard errors for statistics produced from CHIS 2003 data. One

technique for calculating valid standard errors from a complex sample design is replication. For replication, a set of replicate weights needed for the computation have also been included in the



CALIFORNIA HEALTH INTERVIEW SURVEY
UCLA CENTER FOR HEALTH POLICY RESEARCH
10911 WEYBURN AVENUE, SUITE 300
LOS ANGELES, CA 90024
CALL TOLL-FREE: 1-866-275-CHIS (1-866-275-2447)
FAX: (310) 794-2686
WEBSITE: www.chis.ucla.edu

data sets. Some software, such as SUDAAN and WesVar, is specially designed to use these variables in their computations of standard errors. In other packages, such as SAS, there are specific procedures that can accommodate these variables. The replicate weight variables are named as follows:

Adult file— **RAKEDW1** through **RAKEDW80**

Child file—**RAKEDW1** through **RAKEDW80**

Adolescent file—**RAKEDW1** through **RAKEDW80**

For more information on how to use weights with the CHIS data files, please see *How to Use the CHIS 2003 Sample Weights* available on the CHIS Web site at <http://www.chis.ucla.edu/methods.html>.