

# California Health Interview Survey



california  
health  
interview  
survey

Making  
California's  
Voices Heard  
on Health

California Health  
Interview Survey  
UCLA Center for Health  
Policy Research  
10960 Wilshire Blvd.  
Suite 1550  
Los Angeles,  
California 90024

Tel (310) 794-0925  
Toll Free (866) 275-2447  
Fax (310) 794-2686

Email: [chis@ucla.edu](mailto:chis@ucla.edu)

[www.chis.ucla.edu](http://www.chis.ucla.edu)

## Weighting and Estimation of Variance in the CHIS Public Use Files

The California Health Interview Survey (CHIS) Public Use Files (PUFs, separate files for the Adult, Child, and Adolescent samples) each contain 81 different weight variables: RAKEDW0, RAKEDW1, ..., RAKEDW80. These variables fulfill different functions. RAKEDW0 is the final weight that accounts for sample selection probabilities and statistical adjustments for possible undercoverage and nonresponse biases. When this variable is applied in the analysis, it ensures that estimates from the CHIS sample are unbiased of the California population. The remaining weights, RAKEDW1, ..., RAKEDW80, are replicate weights specially designed for valid variance estimation in the absence of confidential person-level sample design information in the CHIS PUFs. The remainder of this document describes how to use these variables to analyze CHIS data.

### Survey Estimates

By applying RAKEDW0, estimates from the survey sample will be representative of the California population (adults, children, or adolescents). If the weight variable is not applied, estimates produced from the sample will represent only the sample and not the population. In order to derive accurate population estimates with CHIS data, simply include the final sample weight, RAKEDW0, in the weight statement. In SAS, for example, it can be stated as:

```
WEIGHT RAKEDW0;
```

Although RAKEDW0 provides estimates at the population level, it does not provide an accurate measure of the variance of an estimate. To properly determine the magnitude of the variability and efficiency of a survey estimate, researchers must take additional steps to

compute the variance.

### Variance Estimation

CHIS employs a multi-stage, geographically stratified random-digit-dial (RDD) sample design. The sampling and estimation techniques used in CHIS are complex and, as a result, the variance of the estimates should not be computed assuming that the data are from a simple random sample. Commonly used statistical packages, such as SAS® (in a limited number of procedures for survey data analysis) and SPSS® (with SPSS Complex Samples™ module), allow valid variance estimation if detailed sample design information, such as sampling strata or cluster, is available in the data file. The CHIS PUFs, however, do not include strata information in order to protect data confidentiality and respondent privacy.

As an alternative, the CHIS PUFs includes replicate weights that incorporate the sample weights and account for the complex sample design. Applying replicate weights allows accurate and efficient variance calculation. The replicate weight variables in the CHIS PUFs are named RAKEDW1, ..., RAKEDW80. These 80 different weights provide variance estimates computed with 80 replications.

WesVar®, SUDAAN® and STATA® are software packages that allow variance estimation using replicate weights. WesVar® uses replication methods such as Jackknife methods (JK), balanced repeated replication (BRR), and Fay's BRR variant. SUDAAN® Version 7.5 onward also provides BRR and JK replication methods. STATA® Version 9 supports JK and BRR. Examples of variance estimation for simple descriptive statistics and cross tabulations using SUDAAN® and STATA® are presented below. Due to limited circulation and use, WesVar® is not discussed further.

*Continued*

### Using Replicate Weights in SUDAAN®

PROC CROSSTAB in SUDAAN® produces weighted frequency and percentage distributions as well as their variances for one-way or multi-way tabulations. It is primarily used for analyses of categorical variables. PROC DESCRIPT produces descriptive statistics for

continuous variables as well as categorical variables. The statistics it produces include total, mean, proportion, geometric mean, and quintile.

For both CROSSTAB and DESCRIPT, the replication method can be applied with the JACKKNIFE design option and the ADJJACK=1 option on the REPWGT statement.

#### Example 1: PROC CROSSTAB

To request one-way table of gender and two-way tables of gender by race/ethnicity [UCLA CHPR classifications]

```
PROC CROSSTAB DATA= data name FILETYPE=SAS DESIGN=JACKKNIFE;
WEIGHT RAKEDW0;
JACKWGTS RAKEDW1--RAKEDW80/ADJJACK=1;
TABLES SRSEX SRSEX*RACEHPR;
SUBGROUP SRSEX RACEHPR;
LEVELS      2      7 ;
RFORMAT SRSEX SRSEXF;
RFORMAT RACEHPR RACEHPF;
RTITLE "SUDAAN USING REPLICATE WEIGHTS FOR JACKKNIFE WITH ADJUSTED FACTOR" "FOR
ROW PERCENTAGE";
RUN;
```

#### Example 2: PROC DESCRIPT

To request distribution of age by gender and race/ethnicity [UCLA CHPR classifications]

```
PROC DESCRIPT DATA= data name FILETYPE=SAS DESIGN=JACKKNIFE;
WEIGHT RAKEDW0;
JACKWGTS RAKEDW1--RAKEDW80/ADJJACK=1;
VAR SRAGE_P;
TABLES SRSEX*RACEHPR;
SUBGROUP SRSEX RACEHPR;
LEVELS      2      7 ;
RFORMAT SRSEX SRSEXF;
RFORMAT RACEHPR RACEHPF;
RTITLE "SUDAAN USING REPLICATE WEIGHTS FOR JACKKNIFE WITH ADJUSTED FACTOR"
"DESCRIPTIVE ESTIMATES FOR SRAGE";
RUN;
```

### Using Replicate Weights in STATA®

STATA® Version 9 includes features enabling replicate variance estimation for survey data using JK and BRR. SVYSET STATA® Version 9 allows specification of

replicate variance estimation methods. Specifically JK variance estimation requires the final sample weight (PWEIGHT) and the replicate weights (JKRW) statement with VCE(JACK) MSE option.

**EXAMPLE 3: SVYSET, SVY: TABULATE**

To request one-way table of gender and two-way tables of gender by race/ethnicity [UCLA CHPR classifications]

```
SVYSET {PWEIGHT=RAKEDW0},
      JKRW(RAKEDW1- RAKEDW80, MULTIPLIER(1)) VCE(JACK) MSE
```

```
SVY: TABULATE SRSEX
```

```
SVY: TABULATE SRSEX RACEHPR
```

**EXAMPLE 4: SVYSET, SVY:MEAN**

To request distribution of age by gender and race/ethnicity [UCLA CHPR classifications]

```
SVYSET {PWEIGHT=RAKEDW0},
      JKRW(RAKEDW1- RAKEDW80, MULTIPLIER(1)) VCE(JACK) MSE
```

```
SVY: MEANS SRAGE_P, OVER(SRSEX RACEHPR)
```

Earlier versions of STATA® do not provide this option, but some researchers have successfully used add-on modules to replication variance estimation for STATA® Version 8. CHIS does not endorse or promote any specific add-on modules; however, such add-on applications are freely available on the web for downloading. Below

are the examples of estimating variance using STATA® SVR add-on module. The replication method can be applied by using the jackknife method (METH JK2) and identifying the final sample weight (PW) and replicate weights (RW).

**EXAMPLE 5: SVRSET, SVRTAB**

To request one-way table of gender and two-way tables of gender by race/ethnicity [UCLA CHPR classifications]

```
SVRSET SET METH JK2
SVRSET SET PW RAKEDW0 RW "RAKEDW1-RAKEDW80"
SVRSET SET DOF 80
```

```
SVRTAB SRSEX
```

```
SVRTAB SRSEX RACEHPR
```

**EXAMPLE 6: SVRSET, SVRMEAN**

To request distribution of age by gender and race/ethnicity [UCLA CHPR classifications]

```
SVRSET SET METH JK2
SVRSET SET PW RAKEDW0 RW "RAKEDW1-RAKEDW80"
SVRSET SET DOF 80
```

```
SVRMEAN SRAGE, BY(SRSEX RACEHPR)
```

**Using Replicate Weights With Other Statistical Software**

As of September, 2005, SAS® and SPSS® do not contain features that allow use of replicate weights to

adjust for survey design in variance estimation. There are other possibilities, like R for utilizing replicate weights, but these are not as commonly used as other software.