

USER'S DIRECTIONS — HINTS 3 (2007) DATA (SPSS)

CONTENTS OF THE ZIP FILE

You will find the following documents included in the zip file:

- 1) READ THIS FIRST—User's Directions: This document—see below for an overview of HINTS 2007, information for downloading and formatting the data, analytic recommendations and sample code to do analyses in Appendices A and B;
- 2) Codebook.HINTS2007: A codebook listing all HINTS 3 (2007) variable names, labels, value labels, inclusion criteria and simple distributions including unweighted and weighted frequencies and percentages;
- 3) Hints 2007-public 2-12-09.sav: SPSS data file;
- 4) Contents.HINTS2007: An alphabetical listing of all HINTS 3 (2007) variables, formats, and variable labels;
- 5) HINTS 2007 Mode Effects-David Cantor: A power point presentation done by David Cantor from Westat (a government contractor) that provides an overview of the dual modes used in HINTS 2007 and recommendations for which mode to use depending on the type of analysis;
- 6) Analytic Changes to Examine Changes Across Years Using HINTS 2003 and 2005 Data: A report that describes how one can perform analyses using combined HINTS data. Some of the procedures found in this report are used for trend analyses and can also be used to test for mode effects in HINTS 3 (2007) (see Appendix A of this document for more information).

OVERVIEW OF HINTS 3 (2007)

HINTS 3 (2007) used a dual mode design where approximately half of the sample was collected through a telephone using random digit dial (RDD) and half completed it through the mail as a pencil and paper questionnaire. Here is the sample size by mode:

RDD completes: 3,767

RDDI partial completes: 325

Total RDD sample size: 4,092

Partial completes were defined as cases where the respondent completed the first section (Health Communications) of the interview, but that did not reach the end of the survey instrument.

Mail completes: 3,473

Mail partial completes: 109

Total Mail sample: 3,582

A questionnaire was considered to be complete if at least 80% of sections A, B, and C were filled in. To be considered partially complete, a questionnaire had to have between 50% and 79% of these sections filled.

HINTS 3 2007 data were collected between January, 2008 and April, 2008. See the HINTS 3 2007 Final Report—found on the HINTS website—for information about response rates.

ANALYZING HINTS DATA

If you are solely interested in calculating point estimates (means, proportions etc.), either weighted or unweighted, you can use programs like SAS, SPSS, STATA and Systat. If you plan on doing inferential statistical testing using the data (i.e., anything that involves calculating a p value or confidence interval), it is important that you utilize a statistical program that can incorporate the replicate weights that are included in the HINTS database. The issue is that the standard errors in your analyses will probably be underestimated if you don't incorporate the jackknife replicate weights; therefore, your p-values will be smaller than they "should" be, your tests will be more liberal, and you are more likely to make a type I error. Statistical programs like SUDAAN, STATA, SAS and Wesvar can incorporate the replicate weights.

At this time, SPSS does not have the capability to incorporate the replicate weights. The data are given in SPSS form for convenience.

Note that analyses of large HINTS domains usually produce reliable estimates, but analyses of small domains may yield unreliable estimates, as indicated by their large variances. The analyst should pay particular attention to the standard error and coefficient of variation (relative standard error) for estimates of means, proportions, and totals, and the analyst should report these when writing up results. It is important that the analyst realizes that small sample sizes for particular analyses will tend to result in unstable estimates.

DUAL FRAME DESIGN

HINTS 3 (2007) is unique from previous iterations in that it utilized a split frame where approximately half of the sample completed the survey by telephone through random digit dial (RDD) and half completed it through the mail as a paper and pencil questionnaire (see the HINTS 3 (2007) Final Report found on the HINTS website for more information). Thus, one can do an analysis with only the RDD respondents, only the mail respondents or both sets of respondents. For each type of analysis you will need to supply the proper final weight (to get population estimates) and replicate weights (to calculate the correct standard errors)-- see Table 1 below for the corresponding names of these variables. See Appendices A and B (attached to this document) for recommendations for using the weights depending on the type of analysis, and sample SUDAAN and STATA code that corresponds to each type of analysis.

MODEFLAG vs. SAMPFLAG

You'll see that there are two variables in the HINTS 3 (2007) data named MODEFLAG and SAMPFLAG. MODEFLAG codes for type of modality used to interview each respondent (i.e., mail or telephone) and the sample sizes correspond to those stated above. SAMPFLAG codes for the sample in which each respondent should be used for analyses that involve mode effects (see Appendix A and B attached to this document). According to frequencies of SAMPFLAG you'll see that there are 11 more subjects in the mail (address) sample as compared to MODEFLAG; this is due to the fact that 11 respondents who were part of the mail sample chose to answer the questions in Spanish and so were interviewed over the telephone (there was no Spanish version of the mail survey). Thus, even though they were interviewed over the telephone, for analytic purposes involving mode, they are considered part of the mail (address) sample.

Table 1 Final Sample Weight and Replicate Weight Variable Names

Survey Mode	Final Sample Weight	Replicate Weights
RDD Only	rwgt0	rwgt1 thru rwgt50
Mail Only	mwgt0	mwgt1 thru mwgt50
Combined RDD and Mail	cwgt0	cwgt1 thru cwgt50

Appendix A

Recommended Use of Weights by Type of Analysis and Survey Mode

1) Descriptive analysis

Example: A researcher wants to estimate a population-level point estimate (e.g., mean, proportion):

- a) For just the sample that completed the mail survey: Use the mail sample weights.
- b) For just the sample that completed the RDD survey: Use the RDD sample weights. Note that by using these weights the resulting population estimate will estimate the result of a telephone census of all households—both those that have landlines and those that are cell-phone only.
- c) Across modes. You may want to first test if there is a difference in population estimates by mode. (See answer to #3, below.) If the difference is not significant, use the weights for the combined sample to calculate a more precise estimate. If there is a significant difference in estimates by mode, report the mode specific estimates and explain if you believe one of the estimates is more accurate than the other estimate or see #4 below for alternative options.

See Appendix B for SUDAAN and STATA code to perform different analyses.

2) Testing for significant relationships between variables (e.g., linear or logistic regression).

Example: A researcher wants to identify (at the population level) significant predictors of a continuous outcome using both categorical (e.g., gender) and continuous (e.g., average minutes of physical activity per week) predictors:

- a) For just the sample that completed the mail survey: Use the mail sample weights.
- b) For just the sample that completed the RDD survey: Use the RDD weights. Note that using the RDD-sample weights will estimate the result of calculating the model parameters from a telephone census of all households—both those that have landlines and those that are cell-phone only
- c) Regardless of mode. You may want to first test if there is a difference in population estimates by mode (see answer to #3 below). If the difference is not significant, use the combined weights to estimate the parameters in the model.

See Appendix B for SUDAAN and STATA code to perform different analyses.

3) Testing for Mode Effects in estimating population parameters

Example: A researcher wants to test if there are mode differences in population estimates of the percentage of people who trust health information from a doctor.

To test for mode effects in estimated population parameters, use a similar procedure for creating weights as described in Analytic Methods to Examine Changes Across Years Using HINTS 1 (2003) and HINTS 2 (2005) Data (included in the data download, see page 8--**see Appendix B for code to create weights and perform analyses**). This will result in creating one final sample weight and 100 replicate weights based on the respective RDD and mail weights. These new weights can be used to determine if the difference of two estimated population parameters is significantly different from zero.

4) Controlling for Mode Effects in estimating model parameters

As a follow-up to #3 above, a researcher found significant mode differences and now wants to develop a regression model to predict the percentage of people who trust health information from a doctor:

- a) Simple regression: Use the new weights created in #3 above and include the SAMPFLAG variable to distinguish between modes.
- b) Multi-variable regression (e.g., including other predictors of interest such as gender or education): Use the new weights created in #3 above and include SAMPFLAG and other variables in your model.
- c) Testing for potential mode by covariate interactions: Use the new weights created in #3 above and include interaction terms between mode (as defined by SAMPFLAG) and other predictors of interest.

5) Testing for trends:

- a) Testing for trends across two iterations of HINTS;

To be consistent with previous iterations of HINTS that used an RDD frame, it is recommended that any trends analyses that include HINTS 3 (2007) data use the RDD weights. To create the appropriate final sample and replicate weights for analyses, use the procedure explained in the report Analytic Methods to Examine Changes Across Years Using HINTS 1 (2003) and HINTS 2 (2005) Data (included in the data download; see page 8.)

- b) Testing for trends using HINTS 2003, 2005, and 2007 data;

To be consistent with previous iterations of HINTS that used an RDD frame, it is recommended that any trends analyses that include HINTS 3 (2007) data use the RDD weights. To create the

appropriate final sample and replicate weights for analyses across all three iterations of HINTS, use the procedure explained in the report [Analytic Methods to Examine Changes Across Years Using HINTS 1 \(2003\) and HINTS 2 \(2005\) Data](#) (included in the data download; see page 8) but extrapolated for three sets of data. Thus, once the complete set of weights is created, there will be one final sample weight and 150 replicate weights. See below:

Final Sample and Replicate Weights for Analyzing Trends Across HINTS 2003, 2005, and 2007:

	Final Sample Weights	Replicate Weights 1-50	Replicate Weights 51-100	Replicate Weights 101-150
HINTS 1 (2003)	2003 Final Weight (fwgt)	2003 Replicate Weights (fwgt1-fwgt50)	2003 Final Weight (fwgt)	2003 Final Weight (fwgt)
HINTS 2 (2005)	2005 Final Weight (fwgt)	2005 Final Weight (fwgt)	2005 Replicate Weights (fwgt1-fwgt50)	2005 Final Weight (fwgt)
HINTS 3 (2007)	2007 Final Weight (rwgt0)	2007 Final Weight (rwgt0)	2007 Final Weight (rwgt0)	2007 Replicate Weights (rwgt1-rwgt50)

Appendix B

Sample SAS, SUDAAN, and STATA Code to Perform Analyses with HINTS 3 (2007)

SAMPLE Code for Use with SUDAAN

1) Use the following code in SUDAAN procedures to specify the sample design:

For RDD only:

```
proc procedurename data=datasetname design=jackknife;
  weight rwtg0;
  jackwgt rwtg1-rwtg50 / adjjack=.98;
```

For Mail only:

```
proc procedurename data=datasetname design=jackknife;
  weight mwgt0;
  jackwgt mwgt1-mwgt50 / adjjack=.98;
```

For the combined data:

```
proc procedurename data=datasetname design=jackknife;
  weight cwgt0;
  jackwgt cwgt1-cwgt50 / adjjack=.98;
```

2a) Use the following SAS code to create one final sample weight and 100 replicate weights to test for mode effects:

```
/*Create Replicate Weights for mode tests involving SAMPFLAG*/

**Replicate Weights;
array mailWgts[50] mwgt1-mwgt50;
array catiWgts[50] rwtg1-rwtg50;
array newWghts[100] nwgt1-nwgt100;

**Adjust Final And Replicate Weights;
if sampFlag eq 1 then do i=1 to 50;
  nwgt0=mwgt0;
  newWghts[i]=mailWgts[i];
  newWghts[50+i]=mwgt0;
end;
else if sampFlag eq 2 then do i=1 to 50;
  nwgt0=rwtg0;
  newWghts[50+i]=catiWgts[i];
  newWghts[i]=rwtg0;
end;
```

2b) Example SUDAAN code to test for mode differences (proc descript):

```
proc descript data=hints design=jackknife;
  weight nwgt0;
```

```

jackwgt1-nwgt100 / adjjack=0.98;
class SampFlag;
diffvar SampFlag=(1 2);
var varname;
print nsum mean lowmean="Lower 95% CI" upmean="Upper 95% CI" semean
t_mean p_mean;
run;

```

SAMPLE Code for Use with STATA

3) Use the following code in STATA procedures to specify the sample design:

For RDD only:

```
svyset [pw=rwgt0], jkrw(rwgt1-rwgt50, multiplier(0.98)) vce(jack) mse
```

For Mail only:

```
svyset [pw=mwgt0], jkrw(mwgt1-mwgt50, multiplier(0.98)) vce(jack) mse
```

For the combined data:

```
svyset [pw=cwgt0], jkrw(cwgt1-cwgt50, multiplier(0.98)) vce(jack) mse
```

4a) Use the following code to create one final sample weight and 100 replicate weights to test for mode effects:

* Adjust final and replicate weights

```
generate nwgt0 = .
```

```
replace nwgt0 = mwgt0 if sampflag == 1
```

```
replace nwgt0 = rwgt0 if sampflag == 2
```

```
foreach i of numlist 1/50 {
```

```
    generate nwgt`i' = .
```

```
    replace nwgt`i' = mwgt`i' if sampflag == 1
```

```
    replace nwgt`i' = rwgt0 if sampflag == 2
```

```
}
```

```
foreach i of numlist 51/100 {
```

```
    local j = `i' - 50
```

```
    generate nwgt`i' = .
```

```
    replace nwgt`i' = mwgt0 if sampflag == 1
```

```
    replace nwgt`i' = rwgt`j' if sampflag == 2
```

```
}
```

4b) Example STATA code to test for mode differences (tabulate procedure):

```
svyset [pw=nwgt0], jkrw(nwgt1-nwgt100, multiplier(0.98)) vce(jack) mse  
svy: tabulate sampflag hc08seekcancerinfo, row se ci format(%8.5f) percent wald
```