

## **Using Ultimate Cluster Models with NAMCS and NHAMCS Public Use Files**

### **I. Background**

Masked sample design variables were included for the first time on NAMCS and NHAMCS public use data files for survey year 2000. These design variables reflected the complex multi-stage sample design of the surveys and were intended for use with software such as SUDAAN that required such data for variance estimation. Following that release, NAMCS and NHAMCS public use files for 1993-1999 were re-released with masked design variables added.

However, difficulties in using the masked design variables soon arose for researchers running software such as SAS, Stata, and SPSS, which all assume a single stage of sampling (ultimate cluster design) for variance estimation. These researchers were unable to use the multi-level design variables as they appeared on the public use files. In addition, further problems occurred when single case strata were found on some of the files, causing some software programs to terminate unsuccessfully.

### **II. New Research**

Research was conducted comparing variance estimation for NAMCS and NHAMCS public use file data using different techniques, including SUDAAN's with-replacement option, SUDAAN's without-replacement option, generalized variance functions, and SAS PROC SURVEYMEANS. Multi-stage design variables were used to develop two new variables, CSTRATM and CPSUM, which could be used with analysis software employing an ultimate cluster design for estimating variance.

The variances produced with these methods were compared with standard errors obtained for in-house files (which contain non-masked design variables), using SUDAAN's without-replacement (WOR) option. This option takes into account the multiple sampling stages of the surveys.

Results showed that use of the masked design variables with the three software applications yielded more accurate standard error estimates than those derived using the generalized variance functions. Standard errors obtained using both full design SUDAAN and the two ultimate cluster designs with masked survey design variables tended to slightly overstate in-house standard errors, on average. This tendency resulted in conservative tests of significance for the data analyzed in the study.

The results support the adoption of the new CSTRATM and CPSUM variables for variance estimation in general, as they were found to yield acceptable results

and can be used with a wide variety of software. Specific findings from this study were published in the journal *Inquiry*<sup>1</sup> and can be obtained from the Ambulatory Care Statistics Branch.

### **III. Implications and Solutions for Data Users**

As a result of this research, the new design variables, CSTRATM and CPSUM, were included on the NAMCS and NHAMCS public use data files for 2002, along with the full set of masked design variables available on public use data files beginning with 1993. However, it is expected that, beginning with survey year 2003, CSTRATM and CPSUM will replace the multi-stage design variables on the public use files.

#### **a. Creating CSTRATM and CPSUM for Data Years Prior to 2002**

Public use data files from 1993-2001 do not contain CSTRATM and CPSUM, but they do contain the information necessary to create them. Until such time as we may re-release older data files with these variables, researchers can create them using the following code:

```
CSTRATM=STRATM;  
CPSUM=PSUM;  
IF CPSUM IN (1 2 3 4) THEN DO;  
  CSTRATM=(STRATM*100000)+(1000*(MOD(YEAR,100)))+  
  (SUBFILE*100)+PROSTRAT;  
  CPSUM=PROVIDER+100000;  
  END;  
ELSE CSTRATM=(STRATM*100000);
```

Examples of using CSTRATM and CPSUM are shown at the end of this section for SUDAAN's 1-stage WR option (with replacement), Stata, SAS PROC SURVEYMEANS, and SPSS.

#### **b. Recoding to Eliminate Single Case Strata (Singletons)**

Since the ultimate cluster procedures discussed above compute Taylor series variance estimates, results should be identical. Results differ, however, when a single case stratum, or singleton, is present on the data file because each software package treats such cases differently. SUDAAN provides an automatic fix-up and generates a warning message when singletons are encountered. However, other software packages may either drop the case or abort the program.

There are no singletons present in the 2001 NAMCS or the 2002 NAMCS and NHAMCS public use files. However, there is a singleton in the 1996 NAMCS public use file, as well as in the 1999 and 2000 NAMCS files.

Singletons also occur in the 1993, 1997, 1998, and 2001 NHAMCS files. We are advising data users to handle these occurrences in the following manner:

### 1. To Eliminate Singletons in Any Year of NHAMCS Data...

Add a line of code **prior to the creation of CSTRATM and CPSUM** which recodes values of 2 to values of 1 for the variable PROSTRAT. An example using SAS syntax is shown below:

```
if subfile = 2 then prostrat = 1;
```

### 2. To Eliminate Singletons in NAMCS Data...

Recoding for NAMCS singletons is year-specific and must be done **after creating CPSUM and CSTRATM** (examples are shown using SAS syntax):

a. For **1996 NAMCS** data, add this code after creating CPSUM and CSTRATM:

```
if subfile=1 and prostrat=14 and year=1996 and stratm=202  
then cstratm=  
203*100000)+(1000*(mod(year,100)))+(subfile*100)+prostrat;
```

b. For **1999 NAMCS** data, add this code after creating CPSUM and CSTRATM:

```
if subfile=1 and prostrat=12 and year=1999 and stratm = 202 then  
cstratm =  
(203*100000)+(1000*(mod(year,100)))+(subfile*100)+prostrat;
```

c. For **2000 NAMCS** data, add this code after creating CPSUM and CSTRATM:

```
if subfile=1 and prostrat=1 and year=2000 and stratm=203 then  
cstratm=(202*100000)+(1000*(mod(year,100)))+(subfile*100)+  
prostrat;
```

### c. Examples Using CSTRATM and CPSUM

Below are examples of how to use CSTRATM and CPSUM with SUDAAN's 1-stage WR (with replacement option), Stata, SAS PROC SURVEYMEANS, and SPSS.

We recommend that NHAMCS data users always read in **both files**, emergency department and outpatient department, when computing variances for emergency and/or outpatient department data in order to **retain the complex sample design and achieve the most accurate results**. If the full set of records from both files is not used for variance estimation, results will tend to understate the true sampling variance. The subpopn statement can be used in SUDAAN to compute results for a subset of records from the full set if desired; consult the documentation for your own software package to find the appropriate method.

#### **SUDAAN 1-stage WR Option**

The program below provides a with-replacement ultimate cluster (1-stage) estimate of standard errors for a cross-tabulation.

```
PROC CROSSTAB DATA=COMB1 DESIGN=WR FILETYPE=SAS;  
  NEST CSTRATM CPSUM/MISSUNIT;
```

#### **Stata**

The pweight (PATWT), strata (CSTRATM), and PSU (CPSUM) are set with the **svyset** command as follows:

```
svyset pweight PATWT  
svyset strata CSTRATM  
svyset psu CPSUM
```

#### **SAS – PROC SURVEYMEANS**

```
PROC SURVEYMEANS DATA=COMB1;  
  CLUSTER CPSUM;  
  STRATA CSTRATM;
```

## **SPSS**

This code pertains to SPSS Inc.'s recently released Complex Samples 12.0 module. It would be used with the "Analysis Preparation Wizard" component of that module. The PLAN FILE statement would be invoked in statistical runs, as in the example for CSTABULATE shown below.

```
CSPLAN ANALYSIS
/PLAN FILE='DIRECTORY\PLANNAME.CSAPLAN'
/PLAN VARS ANALYSISWEIGHT=PATWT
/PRINT PLAN
/DESIGN STAGELABEL= 'ANY LABEL' STRATA=CSTRATM
CLUSTER=CPSUM
/ESTIMATOR TYPE=WR.
```

```
CSTABULATE
/PLAN FILE='DIRECTORY\PLANNAME.CSAPLAN'
/TABLES VARIABLES = var1 var2
/CELLS POPSIZE
/STATISTICS SE
/MISSING SCOPE = TABLE CLASSMISSING = EXCLUDE.
```

For questions regarding the NAMCS and NHAMCS public use files, please contact the Ambulatory Statistics Branch at 301-458-4600, Monday-Friday. 9am-5pm ET.

<sup>1</sup>Hing E, Gousen S, Shimizu I, Burt C. Guide to Using Masked Design Variables to Estimate Standard Errors in Public Use Files of the National Ambulatory Medical Care Survey and the National Hospital Ambulatory Medical Care Survey. *Inquiry*. 40(4):416-415. Winter 2003/2004.

(9/28/2004)