

SUPPLEMENT TO THE 1993-1994 FILE DOCUMENTATION FOR THE NATIONAL AMBULATORY MEDICAL CARE SURVEY (NAMCS) AND THE NATIONAL HOSPITAL AMBULATORY MEDICAL CARE SURVEY (NHAMCS)

I. INTRODUCTION

Since their initial release, NAMCS and NHAMCS public-use files prior to survey year 2000 had never contained sample design information because of confidentiality concerns. This limited the research potential of the data for users who wished to run statistical analyses using software such as SUDAAN, which takes into account the complex sample designs of surveys and thus requires sample design information in order to run properly. It enables the user to compute variances, to perform logistic regressions, and to do other types of sophisticated statistical analysis.

To address this problem, research was conducted at NCHS in methods to include sample design information in the public use files for the NAMCS and the NHAMCS without creating unacceptable disclosure risks. The result of that research was the adoption of a methodology that had as its primary component the collapsing of certain sample strata into larger groups that mask their identity. This made it possible for NCHS to re-release NAMCS and NHAMCS public use files that contain design information. The file names and file layouts are the same as previously published, but the design variables have been appended to the file according to the layout below.

For the release of the 2002 survey data, further research was conducted to address the needs of those using software like Stata, SPSS, and SAS. These products use an ultimate cluster design for variance estimation, which assumes a single stage of sampling, and the multi-stage design variables provided for SUDAAN did not fit those programs. As a result, we have developed two new variables, CSTRATM and CPSUM, which can be used with SUDAAN, SAS, SPSS, and Stata, according to the examples shown at the end of this document.

The 1993 and 1994 NAMCS public use files were re-released previously. This supplemental documentation has been newly updated to address the fact that we are now re-releasing the 1993 and 1994 NHAMCS files, as well as providing updated information on variance estimation using ultimate cluster software. The file layout for the supplemental variables on the public use files is shown in Section III. Corrections and other modifications to the public use data files are described in Section VIII.

II. HOW TO CREATE CSTRATM AND CPSUM

The new design variables, CSTRATM and CPSUM, were included for the first time on the 2002 data file release. If you are using data from surveys prior to 2002, you will need to create CSTRATM and CPSUM using the formula shown below. Beginning with 2003 public use files, it is expected that we will ONLY include CSTRATM and CPSUM on the files and will no longer release the detailed set of masked SUDAAN design variables available prior to that time. Examples of the use of these variables are shown in Section IV.

```
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);
```

III. SUPPLEMENTAL RECORD LAYOUT FOR DESIGN VARIABLES ADDED TO NAMCS/NHAMCS PUBLIC USE FILES FOR 1993-1994

PLEASE NOTE: In addition to the masked design variables, one other item was added to the NHAMCS public use files. This item, PATCODE, which is a sequential numbering of the sample records for each data provider, was omitted from the original 1993-94 NHAMCS releases. It can be found in cols. 366-368 of the emergency department files and in cols. 355-357 of the outpatient department files for both years. The item can be used in conjunction with HOSPCODE to create a unique record identifier.

Item No.	Field Length	File Location	Item Name	Description and Range
1	3	500-502	STRATM	PSU stratum marker (masked) 201-404
2	5	503-507	PSUM	PSU marker (masked) 1-49
3	4	508-511	YEAR	Survey year
4	1	512	SUBFILE	File indicator 1 = NAMCS, 2 = NHAMCS

5	2	513-514	PROSTRAT	Provider stratum marker 1-17 (NAMCS - 1993), 1-15 (NAMCS - 1994), 1-2 (NHAMCS)
6	4	515-518	PROVIDER	Provider code (assigned by ACSB; does not carry across survey years; does not carry across files for hospitals with both ED and OPD records). See ranges below.
7	1	519	DEPT	Hospital dept. code 1 (NAMCS and ED), 2 (OPD)
8	3	520-522	SU	Hospital sampling unit code See ranges below.
9	3	523-525	CLINIC	Hospital clinic code See ranges below.
10	2	526-527	POPPSUM	Stratum counts of PSUs (masked) -1, 1
11	3	528-530	POPSUM	Number of clinic sampling units within a hospital (masked) -1, 0
12	6	531-536	POPVISM	Estimated provider visit volume (masked) -1, 0
13	6	537-542	POPPOVM	Estimated count of in-scope providers within PSU (masked) See ranges below.

PROVIDER ranges:

	NAMCS	NHAMCS-OPD	NHAMCS-ED
1993	6-3405	1-228	1-394
1994	3-3501	1-229	1-397

	NAMCS	NHAMCS-OPD	NHAMCS-ED
SU ranges:			
1993	1	1-70	1-4
1994	1	1-119	1-7
CLINIC ranges:			
1993	1	1-38	1-4
1994	1	1-50	1-4
POPPROVM ranges:			
1993	0-16,832	0-360	0-360
1994	0-19,169	0-360	0-360

IV. USE OF DESIGN VARIABLES

First, please note 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 only a single file is used for variance estimation, results will underestimate the true sampling variance. SUDAAN users can subset their data by means of the SUBPOPN statement where needed; those using other types of software should consult their documentation for the appropriate method.

Below is an example using the multi-level design variables with SUDAAN without-replacement option (WOR) as well as examples using the 1-stage design variables with SUDAAN with-replacement option (WR), SAS, SPSS, and Stata. Please note that based on additional research¹ we recommend the 1-stage design method for variance estimation by those conducting research with NAMCS and NHAMCS public use files.

1. MULTI-LEVEL DESIGN VARIABLES

The variables STRATM, PSUM, YEAR, SUBFILE, PROSTRAT, PROVIDER, DEPT, SU, and CLINIC are all used in the SUDAAN NEST statement, when using the WOR option. The variables POPPSUM, POPPROVM, and POPSUM and POPVISM are used in the TOTCNT statement.

An example of the SUDAAN PROC, NEST, and TOTCNT statements is given below:

```
1 PROC CROSSTAB DATA=COMB1 DESIGN=WOR FILETYPE=SAS;  
2 NEST STRATM PSUM SUBFILE PROSTRAT YEAR PROVIDER DEPT SU  
  CLINIC/MISSUNIT;  
3 TOTCNT POPPSUM _ZERO_ _ZERO_ _ZERO_ POPPROVM _ZERO_  
  POPSUM _ZERO_ POPVISM;
```

2. CSTRATM and CPSUM

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.
```

V. 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.

The 1993 and 1994 NHAMCS emergency department and outpatient department public use files have been examined for occurrences of singletons, and one singleton does appear in the 1993 NHAMCS data. The solution to this is to use the following recode prior to the creation of CSTRATM and CPSUM in your program:

```
IF SUBFILE=2 THEN PROSTRAT=1;
```

This will eliminate singletons in NHAMCS data, but this strategy should not be used for NAMCS data in general. However, no singletons appear in the 1993 or 1994 NAMCS public use files.

For more information on CSTRATM, CPSUM, and singletons, please see the technical paper, "Using Ultimate Cluster Models with NAMCS and NHAMCS Public Use Files," available at the Ambulatory Health Care Data website (www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm). Additional information on guidelines for variance estimation with NAMCS and NHAMCS public use files has been published¹.

VI. SAS INPUT STATEMENTS, VARIABLE LABELS, AND VALUE LABELS

We are also pleased to announce that we have created enhanced documentation for the 1993 and 1994 files for SAS users, consisting of new input statements, format statements, label statements, and format assignments. These files can be downloaded from our Ambulatory Health Care Data website at: <http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm>

VII. SETTING TYPE IDENTIFIER

For data users who wish to combine ambulatory care settings, an identifier for NAMCS, OPD, and ED records can be created using the variables SUBFILE and DEPT as follows:

```
if subfile=1 then settype=1;           /* defines NAMCS records */
if subfile=2 and dept=2 then settype=2; /* defines OPD records   */
if subfile=2 and dept=1 then settype=3; /* defines ED records    */
```

For those using data years 2001 and later, note that we have included the SETTYPE variable on the public use files so you will not have to create it yourself.

VIII. CHANGES TO THE 1993 and 1994 NAMCS AND NHAMCS PUBLIC USE FILES

Some minor revisions have been made to the 1993 and 1994 NAMCS data files as part of the re-release project. This includes corrections to a small number of drug ingredient codes and generic drug codes. These are listed below. (Note that these changes are already reflected on the 1993 and 1994 NAMCS data on SETS CD-ROM.)

Changes to Ingredient Codes:

Old	New
57010	50365
57030	51647
57051	54235
57060	57122

Changes to Generic Codes:

Old	New
56560	57150
57074	56710
57280	52780
92505	56730

In addition, procedure codes were reformatted to replace blanks with dashes in inapplicable fourth digits, for consistency with other years of data. Impute flag data were changed from character format to numeric, by replacing blanks with zeros, which makes the items consistent with the format used in later survey years.

Similar changes were made to the 1993 and 1994 NHAMCS public use data files. On the NHAMCS files, a number of drug therapeutic class codes were also modified to correct missing or erroneous data. Therefore, it is possible that published estimates of drug classes based on earlier data releases will not necessarily match those produced from these revised files.

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.

¹Hing 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)