SUPPLEMENT TO THE 1992 FILE DOCUMENTATION FOR THE NATIONAL HOSPITAL AMBULATORY MEDICAL CARE SURVEY (NHAMCS) OUTPATIENT DEPARTMENT (OPD) AND EMERGENCY DEPARTMENT (ED) PUBLIC USE MICRO-DATA FILES

I. INTRODUCTION

Since their initial release, National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (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 the National Center for Health Statistics (NCHS) to find an appropriate method to include sample design information in NAMCS and NHAMCS public use files 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 with design information. Multi-stage design variables in masked format were released for the first time on the 2000 public use micro-data files, and prior years of data were re-released, back through 1993, to include masked design variables.

Starting with the 2002 survey data, however, 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, NCHS developed two new variables, CSTRATM and CPSUM, which could be used with SUDAAN, SAS, SPSS, and Stata, according to the examples shown at the end of this document. The 2002 NAMCS and NHAMCS public use micro-data files contain both the new and the original versions of the masked design variables. After 2002, only the new version is included.

NHAMCS public use files for 1992 are being re-released to include masked design variables. Although the files are being re-released after the adoption of CSTRATM and CPSUM, the original multi-stage design variables have been included on them for consistency with the earlier file releases.

Researchers who wish to analyze multiple years of data, including years where the new design variables are not on the file (that is, prior to the 2002 NAMCS and NHAMCS), will need to have a consistent set of design variables for all years. CSTRATM and

CPSUM can easily be created for each year of data prior to 2002 that has the original masked design variables by using the code below (Section III). They should be created for each year independently, rather than being created on a file that already combines multiple years, in order to achieve correct results.

In addition to the masked design variables, one other item was added to the 1992 NHAMCS public use files. This item, PATCODE, which is a sequential numbering of the sample records for each data provider, can be used in conjunction with HOSPCODE to create a unique record identifier.

The file names and file layouts are the same as previously published, but the new variables have been appended to the file according to the layout below. One additional change involves the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) procedure codes on the Outpatient Department file, which have been revised to reflect a character format. This format has been in use on the NAMCS and NHAMCS public use files since the 1995 survey release, and the re-released files for 1993 and 1994 included the character formats as well. The original release of the 1992 public use files provided procedure codes in a numeric recode format, where inapplicable third or fourth digits were zerofilled. On the re-released file, inapplicable digits have a dash (-) inserted.

II. NHAMCS OUTPATIENT DEPARTMENT (OPD) AND EMERGENCY DEPARTMENT (ED) RECORD FORMAT FOR SUPPLEMENTAL VARIABLES

ITEM NO.	FIELD LENGTH	FILE LOCATION	ITEM DESCRIPTION
1	3	357-359	[PATCODE] A sequential numbering of all of the records for a hospital. Can be used in conjunction with HOSPCODE to form a unique record identifier. 1-490 (OPD), 1-363 (ED)
2	3	500-502	[STRATM] PSU stratum marker (masked) 201-404
3	5	503-507	[PSUM] PSU marker (masked) 1-49
4	4	508-511	[YEAR] Survey year (1992)

ITEM NO.	FIELD LENGTH	FILE LOCATION	ITEM DESCRIPTION
5	1	512	[SUBFILE] File indicator This has a value of 2 on the NHAMCS file. The item can be used when combining NHAMCS data with data from the National Ambulatory Medical Care Survey. NAMCS files have a value of 1 for this item.
6	2	513	[PROSTRAT] Provider stratum marker. 1, 2
7	4	515	[PROVIDER] Provider code (assigned by the Ambulatory Care Statistics Branch; does not carry across survey years) 1-248 (OPD), 1-407 (ED)
8	1	519	[DEPT] Hospital department code 1 = Emergency department 2 = Outpatient department
9	3	520	[SU] Hospital sample unit stratum marker 1-129 (OPD), 1-6 (ED)
10	3	523	[CLINIC] Hospital clinic code 1-9 (OPD), 1-6 (ED)
11	6	526	[POPPSUM] Stratum counts of PSUs (masked) -1, 1
12	3	528	[POPSUM] Number of clinic sampling units within a hospital (masked). -1, 0

ITEM NO.	FIELD LENGTH	FILE LOCATION	ITEM DESCRIPTION
13	6	531	[POPVISM] Estimated provider visit volume (masked) -1, 0
14	6	537	[POPPROVM] Estimated count of inscope providers within PSU (masked) 0-398

III. 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 the 2003 public use files, CSTRATM and CPSUM are the only design variables included.

```
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);
```

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 research (1) we recommend the 1-stage design method for

variance estimation by those conducting research with NAMCS and NHAMCS public use files. Please keep in mind that multi-level design variables are only on the NAMCS and NHAMCS public use files through the 2002 survey year.

1. SUDAAN WOR Option with Multi-Stage 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.

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

2. 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;
```

3. 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;
```

4. SPSS

This code pertains to SPSS Inc.'s 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 1992 NHAMCS emergency department and outpatient department public use files have been examined for occurrences of singletons, and one singleton does appear in each file. 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.

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 published (1).

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

We are also pleased to announce that we have created enhanced documentation for the 1992 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 1992 NHAMCS PUBLIC USE FILES

Some minor revisions have been made to the 1992 NHAMCS data files as part of the re-release project. This includes corrections to a small number of drug ingredient codes and generic drug codes. Some 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.

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.

For those interested in using drug data, please keep in mind the following information: Drug characteristics are added during data processing for each drug reported on the NAMCS and NHAMCS. The Ambulatory Care Drug Database, the source of this information, is constantly being updated and revised. In order to get the most accurate drug characteristic data, it is recommended that users go to the Ambulatory Health Care Data website and review important information contained in two files there. How you will update the drug information depends on which year of data you are using. Beginning with the 2006 public use file release, NAMCS and NHAMCS switched to a new system based on Multum's Lexicon Plus. If you are planning to trend drug data

across years including 2006, you will need to assign Multum drug characteristics to each drug using a program and data file provided and explained in detail at the AHCD web site.

For those who are only using drug data prior to 2006, it is not necessary to convert all drug characteristics to the Multum version. However, please note that a program can be downloaded from the Ambulatory Health Care Data website that can be used to update the drug characteristics existing on any data file prior to 2005 with the most up-to-date version of the Ambulatory Care Drug Database prior to its conversion to Multum. This means that, for 1992, if the researcher is not using any data after 2005 and so does not require Multum conversion, it is still possible to get the most accurate drug characteristics for the 1992 data, based on the 2005 Ambulatory Care Drug Database. This is explained in more detail at the Ambulatory Health Care website: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/drugs/

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.

(1) 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.

(2/4/2009)