

ICPSR 9212

**National Health Interview Survey,
1979: Smoking Supplement**

*United States Department of Health and
Human Services. Centers for Disease
Control and Prevention. National Center
for Health Statistics*

Technical Documentation

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Terms of Use

The terms of use for this study can be found at:
<http://www.icpsr.umich.edu/cocoon/ICPSR/TERMS/9212.xml>

Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

NOTICE

WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

1979 NATIONAL HEALTH INTERVIEW SURVEY 06/15/2006
PUBLIC USE DATA RELEASE

WARNING - DATA USE RESTRICTIONS! Read Carefully before Using

The Public Health Service Act (Section 308 (d)) provides that the data collected by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), may be used only for the purpose of health statistical reporting and analysis.

Any effort to determine the identity of any reported case is prohibited by this law.

NCHS does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the data files. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users will:

1. Use the data in these data files for statistical reporting and analysis only.
2. Make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery (301-458-4000).
3. Not link these data files with individually identifiable data from other NCHS or non-NCHS data files.

By using these data, you signify your agreement to comply with the above-stated statutorily based requirements.

INTRODUCTION

The National Health Interview Survey (NHIS) is conducted annually by the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC). The NHIS administers face-to-face interviews in a nationally representative sample of households. Each week a probability sample of the civilian non-institutionalized population of the United States is interviewed by personnel of the U.S. Bureau of the Census. Information is obtained about the health and other characteristics of each member of the household.

The NHIS includes a basic questionnaire that remains the same each year plus one or more additional questionnaires that change periodically. More information on the survey in general and the 1979 version, in particular, is found in the Current Estimates

Series report:

http://www.cdc.gov/nchs/data/series/sr_10/sr10_136.pdf

NOTES TO USERS

Users are advised to subscribe to the NHIS Listserv to receive notice of any corrections/updates. For information about joining the Listserv, see the CONTACT INFORMATION section at the bottom of this document.

1979 NHIS FILES AND DOCUMENTATION

The 1979 NHIS files and related documentation have been previously distributed via various formats -- magnetic tape, ASCII CDs, etc.

The files listed below comprise the 1979 National Health Interview Survey (NHIS) public use data release as found on the Internet.

A copy of this file (README.TXT) is located in ASCII format at the following ftp location:

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/1979/

1979 NHIS DATA FILES

Five 1979 NHIS files are referred to as "core" files because they were administered each year (through 1996) without major modification. The five core files are Household, Person, Condition, Doctor Visit, and Hospital.

The 1979 NHIS data have been divided into nine data files, consisting of the five core files and four supplemental files. For ease of transmission, each of the 1979 NHIS data files has been converted into a self-extracting compressed format. To load and expand a compressed file into an ASCII file, perform the following steps:

Download a compressed file onto your hard drive into an appropriate directory (folder). For example: Copy the data file HOUSEHLD.EXE into a directory called NHIS1979. To do this, create a directory on your C: drive called NHIS1979. Then, using your browser, go to the FTP site containing the 1979 NHIS data files (see address at end of this section). Right click on the HOUSEHLD.EXE file. Depending on your browser, select Save Link As... or Save Target As...

Type c:\nhis1979\HOUSEHLD.EXE in the box labeled File Name and hit Enter. To then extract HOUSEHLD.EXE into the ASCII file HOUSEHLD.DAT, go to the directory NHIS1979 and do the following:

- From DOS - in the c:\nhis1979 directory, type HOUSEHLD and hit Enter

- From Windows - Double-click on HOUSEHLD.EXE
 or
 Click the Start Button, Click Run, enter
 c:\nhis1979\HOUSEHLD.EXE and hit Enter.

The ASCII file HOUSEHLD.DAT will be generated which can then be used for processing by PC statistical software packages which utilize ASCII formatted input, or transferred to another computer for processing.

NOTE: The 1979 NHIS files are large in size. Please use caution and check your disk drive capacity before extracting.

1979 NHIS DATA FILES

NHIS FILE TYPE	FILE NAME	LENGTH	# RECS
HOUSEHOLD	HOUSEHLD.DAT	466	41,883
PERSON	PERSONSX.DAT	466	110,530
CONDITION	CONDITON.DAT	466	59,629
HOSPITAL	HOSPITAL.DAT	466	14,606
DOCTOR VISIT	DRVISITX.DAT	466	20,092
EYE CARE	EYECAREX.DAT	507	35,651
HOME CARE	HOMECARE.DAT	588	110,530
RESIDENT MOBIL	RESMOBIL.DAT	494	25,519
SMOKING	SMOKINGX.DAT	373	26,271

Note: The above data files are the unzipped files in ASCII format.

The 1979 NHIS data files are located at:

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/1979

1979 NHIS FILE LAYOUTS (CODEBOOKS)

There is a file layout (also called codebook or data dictionary) for with each data file. The file layouts are in PDF format and can be viewed with Adobe Acrobat software. The Adobe Acrobat Reader software can be downloaded from the Adobe Acrobat Web site at:

<http://www.adobe.com/products/acrobat/readstep2.html>

File Name	Description of Component	Format
EYECAREX.PDF	Eye care layout	Adobe PDF
HOMECARE.PDF	Home Care file layout	Adobe PDF
NHISCORE.PDF*	Core files layout	Adobe PDF
RESMOBIL.PDF	Residential Mobility file layout	Adobe PDF
SMOKINGX.PDF	Smoking file layout	Adobe PDF

*NHISCORE.PDF contains the file layouts for the following five core files: Household, Person, Condition, Doctor Visit, and

Hospital.

The 1979 NHIS data file layout documentation files are located at:

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/1979

1979 NHIS SAS INPUT STATEMENTS

Sample SAS programs containing input statements have been provided for each data file. By using the ASCII data files (.DAT files) as input to these programs, SAS data files can be created. The SAS input statement programs are stored in ASCII format.

File Name	Description of Component	Format
CONDITON.SAS	Condition file SAS program	ASCII
DRVISITX.SAS	Doctor Visit file SAS program	ASCII
HEALTHIN.SAS	Health Insurance file SAS program	ASCII
HEMOCARE.SAS	Home Care file SAS program	ASCII
HOSPITAL.SAS	Hospital file SAS program	ASCII
HOUSEHLD.SAS	Household file SAS program	ASCII
PERSONSX.SAS	Person file SAS program	ASCII
RESMOBIL.SAS	Residential Mobility file SAS program	ASCII
SMOKINGX.SAS	Smoking file SAS program	ASCII

The 1979 NHIS sample SAS input statement programs are located at:

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Program_Code/NHIS/1979

1979 NHIS SURVEY QUESTIONNAIRES AND FLASHCARDS

The 1979 NHIS questionnaires and flashcards can be found as an appendix to the 1979 Current Estimates publication which is on the internet at:

http://www.cdc.gov/nchs/data/series/sr_10/sr10_136.pdf

MISCELLANEOUS

All information collected in the survey is from reports by responsible family members residing in the household. When possible, all adult family members participate in the interview. However, proxy responses are accepted for family members who are not at home and are required for all children and for family members who are physically or mentally incapable of responding for themselves. Although a considerable effort is made to ensure accurate reporting, the information from both proxy respondents and self-respondents may be inaccurate because the respondent is unaware of relevant information, has forgotten it, does not wish to reveal it to an interviewer, or does not understand the intended meaning of a question.

Data from the special health topics (supplements) and core in 19796 were collected for 52 weeks.

The sample for the NHIS is redesigned every decade using population data from the most recent decennial census. The design used in 1979 was used in 1975-1984.

CONTACT INFORMATION

Updates about new data releases, publications, or errors will be sent to members of the NHIS Listserv. To join, visit the CDC website at:

<http://www.cdc.gov/nchs/about/major/nhis/nhislist.htm>

For additional information on the NHIS, visit the website at:

<http://www.cdc.gov/nchs/nhis.htm>

If you have questions or comments on any aspect of the NHIS, please contact us:

E-mail : nhislist@cdc.gov
Call : 301-458-4901
FAX : 301-458-4035
Mail : DHHS:PHS:CDC:NCHS:DHIS
3311 Toledo RD Room 2217
Hyattsville MD 20782
Home Page: <http://www.cdc.gov/nchs/nhis.htm>

For additional information on NCHS data products, contact the:

Data Dissemination Branch, NCHS
3311 Toledo Road
Hyattsville, MD 20782
Tel: 1-301-458-4636 or 1-866-441-NCHS
E-mail: nchsquery@cdc.gov
Internet: <http://www.cdc.gov/nchs/>

STATEMENT OF AUTHENTICITY

This material has been cleared for public distribution by CDC/ATSDR and will be authentic if obtained directly from ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/. CDC/ATSDR takes all effort to assure the authenticity of electronically distributed documents. However, in all instances where the electronic and official agency record differ, the authenticity of the official agency record is controlling.

January 8, 2009

VARIANCE ESTIMATION FOR THE 1973-84 NHIS PUBLIC USE PERSON DATA

Introduction:

This document presents a method for computing standard errors for the 1973-84 NHIS person-level data. It may be used for subsetting data analyses, and it is suitable for analyses of pooled data in the 1973-84 period.

Variance Estimation Method: Single Stage PSUs Sampled With Replacement within Strata Design for 1973-84 NHIS.

NHIS public-use files for 1980-84 contain a three-digit code that is referred to as "Strata and Pseudo Primary Sampling Unit (PSU)" in the file documentation. Equivalent codes were created by NCHS for 1973-1979 but not included in previously-released public-use files. The codes for each year take values 001, 002, ... 298. Each code value denotes a distinct pseudo-PSU for variance estimation, and each pair of code values (e.g., (001,002), (003,004), ... (297,298)) should be grouped together into a pseudo-stratum for variance estimation. That is, for 1973-84 NHIS public-use file variance estimation, there are 298 pseudo-PSUs grouped into 149 pseudo-strata, two pseudo-PSUs per pseudo-stratum.

NCHS has created a variance estimation file for the 1973-84 NHIS public-use files with the following layout:

COLUMNS	FIELD
1-2	Year (e.g., 73, 74, ... 84)
3-5	Random Recode of PSU
6-7	Week
8-9	Segment
10-12	Pseudo-stratum (SUDOSTR)
13-15	Pseudo-PSU (SUDOPSU)

The file is in ASCII format, and it is sorted by Year, Random Recode of PSU, Week, and Segment.

The pseudo-PSU codes for 1980-84 in this file match those previously released by NCHS. Pseudo-stratum numbers 001, 002, ... 149 have been included in this file for the convenience of the user. To assign pseudo-stratum/pseudo-PSU codes to an NHIS file for a given year, sort the NHIS file by Random Recode of PSU, Week, and Segment, and match the 1973-84 NHIS variance estimation file records for that year to the NHIS file. There should be no mismatches.

The NHIS file then should be sorted by pseudo-stratum (SUDOSTR) and pseudo-PSU (SUDOPSU) prior to invoking SUDAAN.

Use the following SUDAAN design statements:

```
PROC <DESCRIPT, CROSSTAB, ...>...    DESIGN = WR;  
NEST SUDOSTR SUDOPSU;  
WEIGHT < weight variable name >;
```

Corresponding statements for other software packages are as follows:

Stata svy:

```
SVYSET [PWEIGHT=< weight variable name >],STRATA(SUDOSTR)PSU(SUDOPSU)  
SVY: MEAN <name of variable to be analyzed for average>  
Or:  
SVY: PROPORTION <name of variable to be analyzed for percentage/proportion>
```

SPSS cdescriptives (for averages) or cstabulate (for percentages/proportions):

One needs first to define a "plan file" with information about the weight and variance estimation, e.g.:

```
CSPLAN ANALYSIS  
/PLAN FILE="< file name >"  
/PLANVARS ANALYSISWEIGHT=< weight variable name >  
/DESIGN STRATA=SUDOSTR CLUSTER=SUDOPSU  
/ESTIMATOR TYPE=WR.
```

And then refer to the plan file when using cdescriptives or cstabulate, e.g.:

```
CSDESCRIPTIVES  
/PLAN FILE="< file name >"  
/SUMMARY VARIABLES =<name of variable to be analyzed>  
/MEAN.
```

```
CSTABULATE  
/PLAN FILE="< file name >"  
/TABLES VARIABLES =<name of variable to be analyzed>  
/CELLS TABLEPCT.
```

SAS proc surveymeans (for averages) or surveyfreq (for percentages/proportions) :

```
PROC SURVEYMEANS;  
STRATA SUDOSTR;  
CLUSTER SUDOPSU;  
WEIGHT < weight variable name >;  
VAR <name of variable to be analyzed>;  
RUN;
```

```

PROC SURVEYFREQ;
STRATA SUDOSTR;
CLUSTER SUDOPSU;
WEIGHT < weight variable name >;
TABLES <name of variable to be analyzed>;
RUN

```

R (including the "survey" package):

(note: R syntax is case-sensitive)

```

# load survey package
require(survey)
# create data frame with NHIS design information, using existing data frame of NHIS data
nhissvy <- svydesign(id=~sudopsu, strata=~sustostr,
                  nest = TRUE,
                  weights=~< weight variable name >,
                  data=< existing data frame name>)
svymean(~<name of variable to be analyzed>,design=nhissvy)

```

note: svymean will produce proportions for "factor variables". Consult the R documentation (<http://cran.r-project.org/manuals.html>) for details.

VPLX:

In the CREATE step, include the following statements:

```

STRATUM    SUDOSTR
CLUSTER    SUDOPSU
WEIGHT     < weight variable name >

```

Then specify the variable to be analyzed in the DISPLAY step:

```

LIST      MEAN(<name of variable to be analyzed>)

```

VPLX can produce percentages by including a CAT statement in the CREATE step. Consult the VPLX documentation (<http://www.census.gov/sdms/www/vdoc.html>) for details.

Subsetted Data Analyses

Frequently, studies of NHIS variables are restricted to select subpopulations, e.g., persons aged 65 and older. To save on storage the user may delete all records outside of the domain of interest. This procedure of keeping only select records is called *subsetting* the data. With a subsetted data set one can produce correct point estimates, e.g., the subpopulation means, but standard errors may be computed incorrectly because some of the sample design information is unavailable to the variance estimation software. **NCHS recommends that subpopulation analyses be carried out using the full data file and the SUBPOPN option in SUDAAN, or an equivalent procedure with another complex design variance estimation software package.**

Subsetting methods with SUDAAN

Strategy 1 (recommended): Use the full data file, and the SUBPOPN statement to identify the subpopulation of interest. For example, if the subpopulation of interest is persons aged 65 and older:

SUBPOPN AGE GE 65 ;

Strategy 2 (not recommended, except when Strategy 1 is infeasible): Use the MISSUNIT option on the NEST statement:

NEST SUDOSTR SUDOPSU/ MISSUNIT ;

In a WR design with exactly 2 PSUs per stratum, when some PSUs are removed from the data file then the SUDAAN MISSUNIT option "fixes" the estimation to avoid errors due to the presence of strata with only one PSU. However, in general there is no guarantee that the variance estimates obtained by this method are equivalent to those obtained using Strategy 1. Other calculations, such as design effects, degrees of freedom, standardization, etc. may need to be carried out differently. The user is responsible for verifying the correctness of their results based on subsetted data.

Implementing Strategy 1 in other software packages can be accomplished as follows:

Stata svy:

Add SUBPOP to the SVY statement, e.g.:

SVY,SUBPOP(AGE>=65): MEAN <name of variable to be analyzed>

SPSS csdescriptives or cstabulate:

One must first define an indicator variable, e.g.:

```
DO IF (AGE GE 65).  
  COMPUTE SUBGRP=1.  
ELSE.  
  COMPUTE SUBGRP=0.  
END IF.
```

And then refer to the indicator variable in csdescriptives or cstabulate, e.g.:

```
CSDESCRIPTIVES (or CSTABULATE)  
/SUBPOP TABLE=SUBGRP
```

It is **very important** that the indicator variable is defined for all data records, otherwise an invalid result can occur.

SAS proc surveymeans or surveyfreq:

One must first define an indicator variable, e.g.:

```
IF AGE >= 65      THEN SUBGRP=1;
                   ELSE SUBGRP=0;
```

And then refer to the indicator variable in proc surveymeans using the DOMAIN statement, e.g.:

```
PROC SURVEYMEANS;
DOMAIN SUBGRP;
```

Proc surveyfreq does not have a DOMAIN statement. Instead, include the indicator variable in the TABLES specification:

```
PROC SURVEYFREQ;
TABLES SUBGRP*<name of variable to be analyzed>;
```

As with SPSS, it is **very important** that the indicator variable is defined for all data records, otherwise an invalid result can occur.

R (including the "survey" package):

After applying the svydesign function to a data frame that contains the entire NHIS sample file being analyzed, create a new data frame using the criteria that define the subgroup of interest. Note that R is very "feisty" when testing for equality, hence the syntax that follows specifies the subgroup of interest without using an equality test.

```
# subset for age>=65 without using equal signs
subgrp <- subset(nhissvy,(age>64))
svymean(~<name of variable to be analyzed>,design=subgrp)
```

VPLX:

In the CREATE step, define one or more CLASS variables that can be used to specify the criteria that define the subgroup of interest.

```
COPY AGE INTO AGECAT
CLASS AGECAT (LOW-64/65-HIGH)
```

The second category of AGECAT defines the subgroup of interest.

Then, specify the variable to be analyzed in the DISPLAY step, and specify the subgroup of interest as well:

```
LIST      MEAN(<name of variable to be analyzed>) /CLASS AGECAT(2)
```

Note that the specification of AGECAT(2) refers to the second category of AGECAT, which is

defined as all values of AGE equal to 65 and all higher values of age that occur in the data.