

About the NSSRN Survey and Data

HRSA's Bureau of Health Workforce has conducted the National Sample Survey of Registered Nurses (NSSRN) approximately every four years since 1977. The data from these periodic surveys provide the basis for evaluating trends and projection of the future supply of nursing resources at the national and state levels.

Available information about the survey and data includes:

- Data Files and Documentation
- Survey Objective
- Survey Design
- About Public Use Files (PUFs)
- Data Documentation
- Data Analysis Formats
- Data Analysis in Microsoft Excel
- Data Files and Documentation

HRSA has made public use data for NSSRN surveys to date available to researchers. For each survey year, HRSA has prepared two Public Use data files in non-delimited ASCII file format. For 2008, SAS-encoded and SPSS-encoded data files are also available.

Public Use data files and documentation for any of the NSSRN surveys may be downloaded from NSSRN Data Download page.

Survey Objective

The objective of the NSSRN is to sample and estimate the characteristics of the registered nurses in the workforce. Nurses may hold licenses in more than one state. Registered nurses participating in the survey answer questions relating to their experience in the field of nursing, including:

- Education and training in nursing
- Professional nursing certifications
- Education and workforce participation prior to becoming a registered nurse
- Current and recent workforce participation
- Income
- Demographic characteristics
- States in which they hold current licenses.

Survey Design

In 2008, the survey design was modified to allow for stratified systematic sampling in each state, with multiple strata developed for age level, dual license, and employment commuting effects. This contrasted with the sample design used from 1977 to 2004 which incorporated a complex, nested sample frame, with equal probabilities of selection of nurses sampled in each state. Probabilities of selection were developed for each record. The samples are selected from current licensure lists in each state. Sampling weights for each state have been calculated and added to the record of each nurse in the respective data files, with adjustments being made in these weights for nurses who have multiple licenses. Though some nurses may be sampled in

sequential surveys, this is a cross-sectional set of survey response files and no attempt is made to track the same nurse's career over time.

About Public Use Files (PUFs)

NSSRN data made available to the public are to be used for research purposes only and may not be used in any manner to identify individual respondents. Most of the respondent information collected from the survey is made available as described below:

State-based Public Use Files – provide information on nurses without identifying the county and metropolitan areas in which they live or work – most users will prefer these files for national or state-level research. Data suppression rules prohibit the publication of information which may allow an individual to derive personally identifiable information about individuals in less-populated areas.

County Public Use Files – provide most, but not all, the same information on the nurse as the State Public Use Files. While the State Public Use Files contain little geographic information below the State level, the County Public Use Files identify the county and metropolitan areas in which the nurses live or work. Data suppression rules prohibit the publication of information which may allow an individual to derive personally identifiable information about individuals in less-populated areas.

The user may not merge the State and County data files into one aggregate database covering all attributes together with extensive geographic information. There are no common, unique identifiers for each surveyed nurse across these two database files.

Data Documentation

For each NSSRN survey cycle and dataset type, there are survey response data and documentation files. These are separated into two complementary sets of zipped files. For ease in downloading, all of the documentation is zipped separately from the data files. In order to keep track of the survey cycle years, the HDW suggests that users maintain the subdirectory names as provided.

The complete documentation includes a detailed PDF file that describes how to use and understand the survey data, as well as copies of SAS and SPSS data description files used for loading the data into SAS or SPSS; some of these files may also be useful once the ASCII data has been loaded into SAS or SPSS. The County documentation file is NSSRN2008_CNTY_Documentation_package.zip (13 MB) and the State documentation file is NSSRN2008_State_Documentation_package.zip (14 MB).

For 2008, the documentation and codebook information are contained in the RN08_State_Documentation.pdf and RN08_CNTY_Documentation.pdf files. This reflects a more streamlined approach to coverage than was published for 1977-2004 where there was a main documentation file, files such as RN04CDOC.pdf and RN04PDOC.pdf, which referenced separate appendix files, respectively, for the County and State documentation. For 1977-2004, accompanying files included the Readme files and Appendices A-I. The appendices within the 2008 State and County Documentation Files roughly correspond to the separate file appendices included from 1977-2004, except as noted below.

Readme files for 1977-2004 are the central listing for summarizing the various files and documentation in each respective zipped directory.

For survey years 1977 to 2004, files such as RN04CDOC.pdf and RN04PDOC.pdf constitute the main documentation manuals for each of the respective General and County public use data sets.

The documentation packages in all years include:

- Background of the survey
- Layout of the documentation manual
- Technical and programmer's information
- Naming conventions for variables in the questionnaire
- Constructed (derived) variables based on formulae using the responses to the original questions of the survey
- Definitions of the derived variables
- Sample variance estimation and design notes, and
- A Codebook, which includes:
 - Documentation identifying locations of each field/variable on the data file
 - Category levels for each field/variable
 - Marginal distribution information for the response categories used in that survey

The appendices cover the following material:

- Appendix A (or appendixa.pdf for 1977-2004) contains a scan of the original questionnaire survey instrument
- Appendix B (or appendixb.pdf for 1977-2004) contains a description of the statistical sampling methodology
- Appendix C (or appendixc.pdf, appCGuid.doc, appCGuid.pdf, appCXwlk.doc, and/or appCXwlk.pdf over 1977-2004), a crosswalk spreadsheet showing the evolution of the various questions by topic, tracks which variables have been available in the State File, the County File, or only in HRSA's in-house file.
- Appendices D through G (or appendixd.pdf to appendixg.pdf over 1977-2004) contain a set of identify the numerical codes for various geographic entities, such as state, foreign country, federal region, county, or metropolitan area. In 2004 only, Appendix H (appendixh.pdf) also provides information on metropolitan areas. Information relating only to the county or metropolitan data is not included in the documentation for the State data files.
- For 2004 only, Appendix I (Appendix I.pdf) provides a list of the priority state orderings which are used in the sampling and weighting processes of the survey. This information is not applicable to the 2008 NSSRN revised systematic stratified sampling design.

The state and county data documentation include tables which crosswalk the various items from the NSSRN surveys against the variable name used in each survey and the respective data set files (i.e., In-House, State Public Use, or County Public Use) in which that variable is located for the respective survey year. This crosswalk is available in Appendix C of the State and County data documentation.

Data Analysis Formats

For the years 1977 to 2004, two pairs of zipped directories are created for each NSSRN year. In the file names below, "xx" represents the last two digits of the year of the survey:

- StatePUFxxdata.zip
- StatePUFxxdoc.zip
- CountyPUFxxdata.zip
- CountyPUFxxdoc.zip

These compressed subdirectories range in size from 3 to 20 MB. Within the documentation package, two SAS and two SPSS syntax auxiliary files are included for all years for each of the respective 1977-2004 NSSRN survey public use file groups. This provides the user ability to generate SAS- and SPSS-encoded data files.

For 2008 only, additional data and format/syntax files may be downloaded. These are contained in the following subdirectories:

- NSSRN2008_State_SAS_encoded_package.zip (31 MB)
- NSSRN2008_CNTY_SAS_encoded_package.zip (32 MB)
- NSSRN2008_State_SPSS_encoded_package.zip (25 MB)
- NSSRN2008_CNTY_SPSS_encoded_package.zip (27 MB)

HRSA provides SAS-encoded and SPSS-encoded data files, as a courtesy to SAS and SPSS users. These zipped files incorporate data files that are encoded and ready to use within the respective SAS or SPSS program, without needing to utilize the supplemental ASCII text data files for 2008. The ASCII formatted data are provided in the following files:

- NSSRN2008_STATE_ASCII_package (18 MB)
- NSSRN2008_CNTY_ASCII_package (20 MB)

The encoded files have much larger file sizes than the ASCII data files. The ASCII data files for 2008 are generally larger than the files from earlier surveys.

The following file names refer to the public use ASCII database files for each survey:

- RNxxPUBL.dat
- RNxxCNTY.dat (1977-2004)
- RN08_State_data.dat
- RN08_CNTY_data.dat (2008)

The text (or.dat) format for the survey response data consists of non-delimited ASCII flat file data records.

For each survey from 1977 to 2008, SAS and SPSS auxiliary syntax files are included with the documentation package. The SAS auxiliary syntax files are in the form of .txt files. The SPSS auxiliary syntax files for 1977-2004 are encoded as .sps files even though they are text files in nature. Each pair of files, for SAS or SPSS respectively, can be used to identify the data file input stream variables on the records, labels for each variable and variable value, and a data format listing for each field found in the ASCII data files.

Both SAS and SPSS users will need to pay heed to the second program line of each of the 'LOADNLABELS' files ('RecFmt' for SAS and SPSS for 2008) which respectively contain an 'infile' statement (SAS) and '/FILE' statement (SPSS) with a default location file name and drive location. The user must substitute their own file name and location for the raw data in ASCII (text) format for each respective public use file. For SPSS in 2008, there is a VarCategories text file which identifies the data value categories for each variable. For 2008 only, there are two

additional files for SAS, a 'format match' file and a format 'sas7bcat' file, in the documentation file grouping; these latter two files individually cover all variables from the survey, including those withheld from the public use data files. These sets of files are to be used in conjunction with the public use data from the survey for the respective SAS or SPSS statistical application program.

Data Analysis in Microsoft Excel

Users may attempt to import the ASCII versions of the database files into Excel; however, because the data files are fixed-length records and are not delimited, caution must be taken in the use of the Excel Import Wizard to ensure proper location of the boundaries between fields.

On top of this consideration, the underlying data contain more columns than some versions of Excel can support. It is necessary to count the number of data columns you have defined, and select and discard unneeded sections of the data record. One way to avoid exceeding the maximum number of columns in an Excel spreadsheet is to mark off one or more blocks of up to 255 characters of the ASCII file as one text field among fields that the user does not intend to analyze further. The import wizard allows the user to skip these blocks of data (using the "Do not import" radio button) in the final step of the import process. Alternatively, the user may elect to import all fields and then subsequently delete each text field which is not of further analytical interest.

In order to make use in Excel of the published weights for each nurse, the user must individually introduce new spreadsheet columns for generation of cross products necessary for obtaining properly-weighted sums and averages.

The HDW believes that users who only possess Excel can successfully perform simple and meaningful analyses of the data if the above steps are undertaken, but users may find manually manipulating the data in Excel a labor-intensive effort. The data warehouse recommends that users employ statistical analysis software such as SPSS, SAS, or Stata to perform complex analyses or compute weighted estimates.