2010 National Survey of Residential Care Facilities

Resident Public-Use Data File

December 2011

Data Description and Usage

Table of Contents

Please Read Carefully Before Working with the Data File

This document describes the data and some of the processes involved in creating the resident public-use data file. It is recommended that a data user reads this document prior to working with the data.

Data Files

The 2010 National Survey of Residential Care Facilities (NSRCF) data are distributed in 2 data files: facility-level and resident-level data. This document refers to the resident public-use file (PUF). The file contains one record for each sampled resident (resident were not interviewed; interviews were conducted with caregivers). This file contains resident characteristics, use of services, charges for care, health status, and cognitive and physical functioning. The facility and resident PUFs cannot be linked using the primary identifier provided in these files. However, the resident PUF contains a limited number of facility–level variables for data users (i.e., SIZE, MSA2, CHAIN, OWN2). This file contains 8,094 records and 220 variables. Each record has a primary identifier (RESNUM). The records in the resident file are sorted in order by the primary identifier.

The data are provided in ASCII format, with fixed-length records. In addition to an ASCII file, separate data files are provided in SAS, STATA, and SPSS formats. PUFs can be downloaded from the NSRCF website as separate files. The individual files for separate download are:

Data files	
SAS	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/SAS_Data/
STATA	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/STATA_Data/
SPSS	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/SPSS_Data/
ASCII	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/Text_ASCII/
Documentation files	
Data dictionary	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NSRCF/2010/2010NSRCF_ResidentsPublicUseFileDataDictionary.pdf
Resident Questionnaire	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Questionnaires/nsrcf/2010/2010_NSRCF_Resident_Questionnaire.pdf
This document (ReadMe file)	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_ResidentsFileDataReadme.pdf
Restricted Variables	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_ResidentsFileRestrictedvariables.pdf

Documentation

There are several types of documentation available for use with the 2010 NSRCF resident data release. These include a data dictionary or codebook, the Resident Questionnaire, a methods report, and this data description or readme file. Information on sampling design and data collection is available elsewhere (http://www.cdc.gov/nchs/data/series/sr_01/sr01_054.pdf).

Data dictionary

The resident data dictionary or code book is provided as a single file containing three major sections in the Resident Questionnaire: A) Background; B) Characteristics; and C) Health Status and Physical Functioning. Each variable in the resident PUF has its own codebook entry. Computer-Assisted Personal Interviewing (CAPI) system used for data collection allowed interviewers to ask only questions specific to the individual facility or resident, skipping unnecessary questions. If a question or a series of questions in the survey were legitimately skipped reflecting question skip pattern, responses to the question were coded as "-1= LEGITIMATE SKIP" in the data dictionary. The question skip pattern is specified in the data dictionary besides the question text and code categories. The data users are advised to consult the questionnaire to better understand these question skip patterns.

Resident Questionnaire

A user-friendly version of Resident Questionnaire is included in the data release package and available at:

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Questionnaires/nsrcf/2010/2010_NSRCF_Resident_Questionnaire.pdf

The questionnaire includes all the questions asked during the facility interview along with the skip pattern of questions. There may be some differences in how questions were asked in the questionnaire, and how they are coded in PUF. Also some questions may not be available as data in the public-use file. These differences are largely related to efforts to reduce disclosure risk. For instance, the PUF may have fewer response categories (response categories collapsed) than the number of categories indicated in the questionnaire.

Data processing activities to create the public use file

The raw data received from the field were reviewed and edited prior to releasing the PUF. Data were reviewed for accuracy, logic, consistency and completeness. Additionally, extensive disclosure risk review was conducted to prevent the identity of any facilities who participated in the survey from being made known to the public. NCHS staff used various methods to perturb the data to minimize disclosure risk, and then ensured that the perturbation did not affect the estimates. Following methods were employed on the restricted, in-house file to create the PUF:

Changes in data because of respondent comments

The interviewer recorded comments made by the respondent during the course of the interview. NCHS staff reviewed these comments for selected questions (e.g., CHARGES). The original answer was changed if it was determined that the comment changed the substance of the recorded answer. For some cases, a change resulting from the comment review process led to some inconsistencies with responses provided for subsequent questions. Therefore, when editing

to resolve inconsistencies in the data, a limited number of changes were made to responses to related questions within the same questionnaire section. No changes were made to responses to questions in other sections, or in the facility data to ensure internal consistency. For example, if the monthly base rate reported for a sampled resident in a room designed for one person in a regular, non-Alzheimer's unit provided in the resident data was different from the average monthly base rate for living in the same type of living quarters in the facility data, no changes were made.

Item nonresponse and imputed data

Item nonresponse is a source of missing data and occured if a respondent did not know the answer to a question or refused to answer a question, the interviewer inadvertently skipped a question due to problems relating to CAPI or the interview broke off before administering the entire questionnaire. In the data file, item-nonresponse is coded as -8 when a respondent did not know the answer to a question ("Don't Know"), as -7 when a respondent refused to answer a question ("Refused"), and as -9 when a question was inadvertently skipped or never asked ("Not Ascertained").

1. NCHS handled item nonresponse for race and ethnicity by imputing nine cases for race and four cases for ethnicity, using a hot-deck method.

2. For selected variables, missing data were imputed. Imputed variables are flagged for the data users. If a variable was imputed (e.g., HISPAN), a flagging variable was included to indicate cases imputed for the variable (e.g., HISPAN_FL).

Masked variables

To protect the confidentiality of the information respondents provided, variables were masked or simply not included in the public-use data. In making these modifications, NCHS staff tried to maintain a balance between the need for data confidentiality and user needs for data.

1. Direct identifiers were not included in the PUF: Names, addresses, geographic information (region, state, metropolitan statistical area status for extra-large facilities), are not included in the PUF.

2. Modified variables:
    a. Ownership (OWN2): This variable was recoded into a two-category variable from a three category variable, by collapsing private nonprofit ownership with state, county, or local government ownership.
    b. Other categorical variables collapsed into fewer categories: Some categorical variables were collapsed into fewer response categories. These included MOVEFROMRC, FSTATRC, PSTATRC, and HEARINGRC.

c. Race of resident (RACERCPUF): American Indians and Alaska Natives, Hawaiian or other Pacific Islanders, and multiple race residents were collapsed into an "other race" category in the public use file.

d. Length of stay in facility (LOSRC): Instead of including the exact date when a sampled resident moved into a facility, a categorical length of stay variable was provided.

3. Top/bottom coded variables
   a. For some variables, upper and lower values were recoded for confidentiality purposes (top- or bottom- coded). Upper or lower cut-off points were used and values above the upper limit or below the lower limit included in the top and bottom categories. These included CHARGES, NUMEMER, and HOSPTREAT.

4. Additional case-specific modifications:
   a. Amenities for residents in small facilities. In small facilities, only a few residents indicated that they had the following amenities: microwave, cook top or hot plate, and oven. In the public use file, these residents in small facilities were recoded as not having these amenities, like the majority of residents in the small facilities: 27 residents for RESMICRO, 11 residents for RESCOOK, and 13 residents for RESOVEN.
   b. Bathroom located inside room or apartment (RESBATH): Only a few residents living in extra-large facilities indicated that they did not have a bathroom in their room or apartment. These residents (20 cases) were recoded as having a bathroom in their room or apartment like the overwhelming majority of residents in extra-large facilities.

Other recoded/derived variables
Several variables were recoded and new variables were derived for ease of analysis.

1. Race-ethnicity: race and Hispanic origin were combined to derive a race-ethnicity variable (RACEETH).

2. Resident vision: Two questions were asked about visual impairments (BLIND, SEEPROB). These variables were combined to provide a summary of a resident's vision (VISION).

3. Resident treated in hospital emergency room and number of times treated: Two questions indicating whether a resident was treated in a hospital emergency room (HOSPTREAT) and the number of times treated there (NUMEMER) were combine to provided a summary of treatment in a hospital emergency room (HOSPTREATRC).

4. Several ADLs and Nagi variables were recoded to fit the skip pattern in the questionnaire. For instance, a separate category "confined to bed or chair" was added to include those cases that were skipped out.

5. Assistance with toileting: A new variable was created to combine assistance using the toilet (TOILHLP) and assistance with ostomy (OSTOMHLP). The derived variable TOILHLPRC provides a more complete picture of the number of residents that receive help with toileting.

6. Total number of ADL limitations: A summary variable TOTHLP was created to indicate the total number of ADLS for which a resident needed assistance. The summary combined EATHLP, DRESSHLP, BATHHLP, TOILHLPRC & TRANSHLPRC. Residents who were confined to bed or chair were skipped out of the transferring question (TRANSHLP). Therefore, TOTHLP varied between 0-5 for residents who were not confined to bed/chair, and 0-4 for residents confined to bed/chair. All individual ADL variables are provided in the PUF so that users may choose to recode differently.

Variables with "other specify" responses

A showcard was used for the question about resident health conditions. Respondents could select from 30 health conditions or groups of health conditions listed on the showcard. In addition, the showcard included a separate category for "other conditions" that were not listed on the card.

1. If a resident had a condition that was not listed on the card, the respondent could specify what the condition was. Responses to this "other specify" question were examined and responses were recoded to existing categories if they matched them.

2. A new category indicating gastrointestinal condition (GI) was also created to code the large number of responses that indicate health conditions associated with the gastrointestinal system.

Additional comments

Some inconsistencies in the data were identified but not corrected because of a lack of valid basis for making a correction. These include the following:

1. There were 49 residents who lived in a dementia unit (INDEMENT) but were not coded as having Alzheimer's disease (ALZ). No changes were made.

2. Some residents who were coded as being confined to bed or chair (INBED=1 or INCHAIR=1) also had interviewer comments indicating that the resident could walk. For instance, there was one resident for whom the interviewer indicated that "the resident

only comes down for dinner." In these cases, no changes were made in the inbound status of residents (INBED/INCHAIR) since the answers to INCHAIR and IBED were based on the perceptions of the respondents.

3. The activity variables (ACTIVITY, OUTSIDE) for 152 and 7041 residents, respectively were coded '0' indicating that these residents did not participate in any of the listed activities. However, the code "0" may not necessarily indicate inability to perform these activities. It may be indicative of personal preferences and other factors in addition to functional ability.

List of restricted variables
A complete list of the resident PUF variables that are masked or restricted is available at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010ResidentsRestrictedvariables.pdf. Users wishing to access these restricted variables or link the resident PUF to the facility PUF need to contact the National Center for Health Statistics (NCHS) Research Data Center (http://www.cdc.gov/rdc/index.htm).

Reliability of estimates

Estimates published by NCHS must meet reliability criteria based on the relative standard error (RSE or coefficient of variation) of the estimate and on the number of sampled records on which the estimate is based. The RSE is a measure of variability and is calculated by dividing the standard error of an estimate by the estimate itself. The result is then converted to a percentage by multiplying by 100. Guidelines used by NCHS authors to determine whether estimates should be presented in tables of NCHS published data reports include the following:

- If the estimate is based on 60 or more sampled cases and the RSE is less than 30%, the estimate is reported and is considered reliable.
- If the estimate is based on fewer than 30 sampled cases, the value of the estimate is not reported. This is usually indicated with an asterisk (*).
- All other reported estimates should not be assumed to be reliable. These include estimates with an RSE of 30% or more and estimates based on 30–59 cases, regardless of RSE.

The data collected in the 2010 NSRCF are obtained through a complex, multistage sample design that involves stratification and clustering. The final weights provided for analytic purposes have been adjusted in several ways to yield valid national estimates for residential care facilities in the U.S. Users are reminded that the use of standard statistical procedures that are based on the assumption that data are generated via simple random sampling (SRS) generally will produce incorrect estimates of variances and standard errors when used to analyze data from the NSRCF. The clustering protocols that are used in the multistage selection of the NSRCF

sample require other analytic procedures, as described below. Users who apply SRS techniques to NSRCF data generally will produce standard error estimates that are, on average, too small, and are likely to produce results that are subject to excessive Type I error.

In this document, examples of SUDAAN computer code are provided for illustrative purposes. Examples are provided for the SAS, STATA, and SPSS software packages. However, the appropriate application of these procedures is the ultimate responsibility of users. NCHS strongly recommends that NSRCF data analyzed under the direction of or in consultation with a statistician who is cognizant of sampling methodologies and techniques for the analysis of complex survey data. The NSRCF resident PUF includes design variables that designate each record's stratum marker and the first-stage unit (or cluster) to which the record belongs. Examples follow for using these design variables with SUDAAN, STATA, and SAS survey procedures.

Table 1a. Computations using SUDAAN

PROC statement	NEST statement	TOTCNT statement	WEIGHT statement
PROC x FILE = y DESIGN = WOR;	NEST RSTRATUM;	TOTCNT RPOPFAC RPOPRES;	WEIGHT RESFNWT;

Table 1b. Computations using STATA

Design description in STATA
svyset facility [pweight=resfnwt], strata(rstratum) fpc(rpopfac) vce(linearized) singleunit(centered) || resident

Table1c. Computations using SAS

PROC	STRATA	CLUSTER	WEIGHT
PROC SURVEY_ DATA = Y TOTAL = SECONDFILE;	STRATA RSTRATUM;	CLUSTER FACILITY;	WEIGHT RESFNWT;

Table 1d. Computations using SPSS
Users need to define a "plan file" with information about the weight and variance estimation. There may be other ways to create a plan file and obtain accurate standard errors. But here we suggest one way to define "a plan file" that produces comparable estimates and standard errors as SUDAAN, SAS and STATA.

1. Select *Analyze → Complex Samples → Prepare for Analysis* from the menu bar.
2. Stage 1: Design Variables
    * Select "RSTRATUM" from the variable list as **Strata**.
    * Select "RESFNWT" as **Sample Weight**.

3. Stage 1: Estimation Method
   - Select "Equal WOR (equal probability sampling without replacement)"
4. Stage 1: Size
   - Select "Population Sizes" as **Units** from the drop down menu.
   - Select "RPOPFAC" for **Read values from variable**
5. Stage 1: Plan Summary
   - Select "Yes" for **Do you want to add stage 2?**
6. Stage 2: Design Variables
   - Select "POPRES" from the variable list as **Strata.**
7. Stage 2: Estimation Method
   - Select "WR (sampling with replacement)"

Log will be created as shown below:

```
CSPLAN ANALYSIS
 /PLAN FILE='<file name>'
 /PLANVARS ANALYSISWEIGHT=RESFNWT
 /PRINT PLAN
 /DESIGN STRATA=RSTRATUM CLUSTER=FACILITY
 /ESTIMATOR TYPE=EQUAL_WOR
 /POPSIZE VARIABLE=RPOPFAC
 /DESIGN STRATA=RPOPRES
 /ESTIMATOR TYPE=WR.
```

CAUTION: Standard errors produced by SPSS Complex Samples tend to be slightly smaller than those produced by other statistical software. Examples of standard errors for GENDER produced by SUDAAN and SPSS are compared:

Standard Error	SUDAAN	SPSS Complex Samples
1=Male	5357.27	5357.08
2=Female	7186.77	7185.67
Total	7486.79	7485.60

Obtaining the data

NSRCF data files are available for free of charge to users and analysts and can be downloaded from the NSRCF web site. There are a few conditions or restrictions for data use and they include:

1. Use the data in this dataset 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.
3. Report apparent errors in the NSRCF data or documentation files to the Long-Term Care Statistics Branch (LTCSB).

We also request the user inform LTCSB of any publications or presentations produced based on the 2010 NSRCF data, and cite relevant NSRCF documentations/ data products in their work when appropriate.

Contact Information

For questions, suggestions, or comments concerning NSRCF data, please contact the LTCSB at:

Long-term Care Statistics Branch (LTCSB), NCHS,

3311 Toledo Road, Hyattsville, MD 20782

E-mail: ltcsbfeedback@cdc.gov

Phone: 301-458-4747.