

2010 National Survey of Residential Care Facilities

Facility Public-Use Data File

December 2011

Data Description and Usage

(Revised 2/15/2012)

Table of Contents

Data Files.....	4
Documentation	4
Data dictionary.....	4
Facility Questionnaire	5
Data processing activities to create the public-use file	5
Consistency checks.....	5
Changes in data because of respondent comments.....	6
Item nonresponse and imputed data	6
Masked variables	7
Variables with “other specify” responses.....	9
List of restricted variables.....	9
Reliability of estimates.....	9
Obtaining the data	11
Contact Information.....	12

Please Read Carefully Before Working with the Data File

The Public Health Service Act (Section 308 (d)) provides that the data collected by 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. 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 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.

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

This document describes the data and some of the processes involved in creating the facility public-use data file. We recommend that a data user read 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 facility file. The facility file contains one record for each sampled facility that provided an interview. This file contains characteristics about facilities, types of staff employed, and policies on admission, retention, and discharge. This file contains 2,302 records and 308 variables. Each record contains a primary identifier (FACILID). The records in the facility file are sorted in the order of 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_FacilityPublicUseFileDataDictionary.pdf
Facility Questionnaire	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Questionnaires/nsrcf/2010/2010_NSRCF_Facility_Questionnaire.pdf
This document (ReadMe file)	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_FacilityFileDataReadme.pdf
Restricted Variables	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_FacilityFileRestrictedvariables.pdf

Documentation

There are several types of documentation available for use with the 2010 NSRCF facility data release. These include a data dictionary or codebook, the Facility Questionnaire, a methods report, a data brief, 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 facility data dictionary or code book is provided as a single file containing all four sections in the Facility Questionnaire: A) Facility Characteristics; B) Facility Policies and Services; C)

Background of Administrator; and D) Demographics of Administrator. Each variable in the 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, skipping unnecessary questions. If a question or a series of questions in the survey were legitimately skipped for selected respondents, responses of these respondents 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 the question skip patterns.

Facility Questionnaire

A user-friendly version of Facility 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_Facility_Questionnaire.pdf

The questionnaire includes all the questions asked during the facility interview along with the skip patterns for selected questions. There may be some differences in how questions were asked in the questionnaire, and how they are coded in the PUF. Also answers to some questions may not be available in the public-use file. These differences are largely related to efforts to reduce disclosure risk. For instance, the PUF may provide percentages for some variables (e.g., percent of residential care rooms or apartments that are rooms designed for one person, percent of residents with some or all of their long-term care services paid by Medicaid), while the questionnaire asked for specific numbers (e.g., number of rooms that are designed for one person, total number of residential care rooms or apartments, number of residents with some or all of their long-term care services paid by Medicaid).

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:

Consistency checks

1. To ensure internal consistency of the data, for some questions, edit checks were programmed into the CAPI system and were applied during data collection. These edits were programmed based on the expected range of responses for given questions and the logical consistency between questions. For instance, the CAPI system prompted

interviewers to verify if the total number of residents that the respondent provided was greater than the total number of beds.

2. In addition, twenty-six facilities were identified as having inconsistencies in the number of rooms, apartments, and bathrooms provided during the in-person interview, and were re-contacted for further clarification of the discrepant data. Some of these facilities confirmed the numbers, while others did not. Certain reporting patterns were identified that provided some explanations for the inconsistencies (e.g., provided the number of beds instead of the number of rooms and apartments, double counting of rooms within apartments, double counting of rooms with a full bathroom).
 - a. If facilities presented similar reporting patterns relative to the inconsistent number of rooms and apartments and bathrooms, the numbers were edited to resolve the inconsistencies.
 - b. For other cases that had different reporting patterns and there were no respondent comments available to suggest alternate responses, the numbers were ratio-adjusted using the total number of rooms or apartments as the gold standard for distributing the total into the various room type components.

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., number of rooms or apartments by type, number of residents that had Medicaid pay some or all of their long-term care services, total number of hours worked by staff type, average monthly base rate by accommodation type, and various services included in base rate or provided at an extra charge). 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 resident data to ensure internal consistency. For example, if average monthly base rate for a room designed for one person in a regular, non-Alzheimer's unit that the facility provided in the facility data was different from the monthly base rate reported for a sampled resident living in the same type of living quarters in the resident data, no changes were made.

Item nonresponse and imputed data

Item nonresponse is a source of missing data, and occurred 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 if 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”). For selected variables, missing data were imputed. Imputed variables are flagged in the data file. If a variable was imputed (e.g., FLUSTAFF), a flagging variable is included to indicate cases imputed for the variable (e.g., FLUST_IMP). Imputations were done using the following methods:

1. For a few variables (i.e., chain status, ownership type), the missing information was entered into the data file from other sources including establishment websites.
2. For a few other selected variables, NCHS handled item nonresponse by imputations. The imputation processes involved converting the -7 (“Refused”) and -8 (“Don’t Know”) to missing, determining control variables that helped to identify which cases should have responses for the imputed variables, and determining an appropriate imputation technique. A variable indicating percentage of facility’s employees receiving a flu shot last flu season (FLUSTAFF) had the most cases (81 cases) with missing data for the facility survey. For the 81 cases with missing data, a mixed method of imputation (combination of regression and hot-deck) was used with the predicted value of the dependent variable in a regression model used as a primary sort for the hot-deck method. The set of covariates for the regression included geographic characteristics, size of facility, ownership type, and chain status.
3. The NSRCF facility file provides summary data on the demographic characteristics of the residents in sample facilities. These include age, gender, race, and ethnicity of residents. When these variables were missing in the data file, NCHS imputed the data by applying the mean proportion in each characteristic group within size and region categories. For instance, if the proportion of residents who are male was missing for a facility, we imputed for this case by entering the proportion of residents who are male in a facility of similar size and in the same region.

Masked variables

To protect the confidentiality of the information respondents provided, a number of variables have been 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 are 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. Some variables have been modified to minimize disclosure risk. For instance, the number of beds in a facility and the number of residents in a facility are not

Please Read Carefully Before Working with the Data File

The Public Health Service Act (Section 308 (d)) provides that the data collected by 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. 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 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.

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

This document describes the data and some of the processes involved in creating the facility public-use data file. We recommend that a data user read 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 facility file. The facility file contains one record for each sampled facility that provided an interview. This file contains characteristics about facilities, types of staff employed, and policies on admission, retention, and discharge. This file contains 2,302 records and 308 variables. Each record contains a primary identifier (FACILID). The records in the facility file are sorted in the order of 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_FacilityPublicUseFileDataDictionary.pdf
Facility Questionnaire	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Questionnaires/nsrcf/2010/2010_NSRCF_Facility_Questionnaire.pdf
This document (ReadMe file)	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_FacilityFileDataReadme.pdf
Restricted Variables	ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_FacilityFileRestrictedvariables.pdf

Documentation

There are several types of documentation available for use with the 2010 NSRCF facility data release. These include a data dictionary or codebook, the Facility Questionnaire, a methods report, a data brief, 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 facility data dictionary or code book is provided as a single file containing all four sections in the Facility Questionnaire: A) Facility Characteristics; B) Facility Policies and Services; C)

Background of Administrator; and D) Demographics of Administrator. Each variable in the 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, skipping unnecessary questions. If a question or a series of questions in the survey were legitimately skipped for selected respondents, responses of these respondents 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 the question skip patterns.

Facility Questionnaire

A user-friendly version of Facility 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_Facility_Questionnaire.pdf

The questionnaire includes all the questions asked during the facility interview along with the skip patterns for selected questions. There may be some differences in how questions were asked in the questionnaire, and how they are coded in the PUF. Also answers to some questions may not be available in the public-use file. These differences are largely related to efforts to reduce disclosure risk. For instance, the PUF may provide percentages for some variables (e.g., percent of residential care rooms or apartments that are rooms designed for one person, percent of residents with some or all of their long-term care services paid by Medicaid), while the questionnaire asked for specific numbers (e.g., number of rooms that are designed for one person, total number of residential care rooms or apartments, number of residents with some or all of their long-term care services paid by Medicaid).

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:

Consistency checks

1. To ensure internal consistency of the data, for some questions, edit checks were programmed into the CAPI system and were applied during data collection. These edits were programmed based on the expected range of responses for given questions and the logical consistency between questions. For instance, the CAPI system prompted

interviewers to verify if the total number of residents that the respondent provided was greater than the total number of beds.

2. In addition, twenty-six facilities were identified as having inconsistencies in the number of rooms, apartments, and bathrooms provided during the in-person interview, and were re-contacted for further clarification of the discrepant data. Some of these facilities confirmed the numbers, while others did not. Certain reporting patterns were identified that provided some explanations for the inconsistencies (e.g., provided the number of beds instead of the number of rooms and apartments, double counting of rooms within apartments, double counting of rooms with a full bathroom).
 - a. If facilities presented similar reporting patterns relative to the inconsistent number of rooms and apartments and bathrooms, the numbers were edited to resolve the inconsistencies.
 - b. For other cases that had different reporting patterns and there were no respondent comments available to suggest alternate responses, the numbers were ratio-adjusted using the total number of rooms or apartments as the gold standard for distributing the total into the various room type components.

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., number of rooms or apartments by type, number of residents that had Medicaid pay some or all of their long-term care services, total number of hours worked by staff type, average monthly base rate by accommodation type, and various services included in base rate or provided at an extra charge). 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 resident data to ensure internal consistency. For example, if average monthly base rate for a room designed for one person in a regular, non-Alzheimer's unit that the facility provided in the facility data was different from the monthly base rate reported for a sampled resident living in the same type of living quarters in the resident data, no changes were made.

Item nonresponse and imputed data

Item nonresponse is a source of missing data, and occurred 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 if 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

provided in the PUF, but are replaced by a 4-category occupancy rate variable (OCCU_CAT).

- b. Similarly, instead of providing the actual number of resident living quarters with different kitchen and other amenities by specific room or apartment type (e.g., number of rooms with a microwave, number of apartments with a door to hallway that can be locked from inside), the PUF includes a set of variables indicating percent of resident living quarters with different kitchen amenities regardless of room or apartment type (e.g., UNITMICRO for percent of rooms or apartments with a microwave, UNITLOCK for percent of rooms or apartments with a door to hallway that can be locked from inside).
 - c. Response categories were collapsed into fewer categories for some variables. For example, a variable on ownership type (OWNERSHP) was recoded to a 2-category variable (OWN2), by collapsing private nonprofit ownership with state, county, or local government ownership.
3. Top or bottom coded variables
- a. For some variables, upper and lower values were recoded for confidentiality purposes (top- or bottom- coded). For example, a variable indicating the percent of residents with some or all of their long-term care services paid by Medicaid during the last 30 days (RESMED2) was bottom coded at 20% for values between 0% and 20%, and top coded at 75% for values greater than or equal to 75%.
 - b. Likewise, a variable indicating the total number of hours worked by registered nurses (RNs) to provide residents direct care in the past 7 days (RNHOUR2) was top coded at 80 and the rest of values from 0 to 79 were kept as continuous. In addition to providing the actual number of direct care hours by staff type, top and bottom coded, variables indicating hours per patient per day by selected staff type (i.e., registered nurses, licensed practical nurses or licensed vocational nurses, personal care aides) were also included, since number of residents is not provided in the PUF.
4. Additional case-specific modifications:
- a. There was only one medium facility with 11 to 25 beds that had 100% of its resident living quarters be two-bedroom or three-bedroom apartments. In order to mask this case, we decided to recode the facility to have some of its resident living quarters also be studio or one-bedroom apartments.
 - b. Similarly, only a few small facilities with 4 to 10 beds had their resident living quarters equipped with a cooktop/hot plate or an oven. In order to mask these cases, they were coded as not having these amenities. There were no statistically significant differences in estimates as a result of this modification. However, this

provided in the PUF, but are replaced by a 4-category occupancy rate variable (OCCU_CAT).

- b. Similarly, instead of providing the actual number of resident living quarters with different kitchen and other amenities by specific room or apartment type (e.g., number of rooms with a microwave, number of apartments with a door to hallway that can be locked from inside), the PUF includes a set of variables indicating percent of resident living quarters with different kitchen amenities regardless of room or apartment type (e.g., UNITMICRO for percent of rooms or apartments with a microwave, UNITLOCK for percent of rooms or apartments with a door to hallway that can be locked from inside).
 - c. Response categories were collapsed into fewer categories for some variables. For example, a variable on ownership type (OWNERSHP) was recoded to a 2-category variable (OWN2), by collapsing private nonprofit ownership with state, county, or local government ownership.
3. Top or bottom coded variables
- a. For some variables, upper and lower values were recoded for confidentiality purposes (top- or bottom- coded). For example, a variable indicating the percent of residents with some or all of their long-term care services paid by Medicaid during the last 30 days (RESMED2) was bottom coded at 20% for values between 0% and 20%, and top coded at 75% for values greater than or equal to 75%.
 - b. Likewise, a variable indicating the total number of hours worked by registered nurses (RNs) to provide residents direct care in the past 7 days (RNHOUR2) was top coded at 80 and the rest of values from 0 to 79 were kept as continuous. In addition to providing the actual number of direct care hours by staff type, top and bottom coded, variables indicating hours per patient per day by selected staff type (i.e., registered nurses, licensed practical nurses or licensed vocational nurses, personal care aides) were also included, since number of residents is not provided in the PUF.
4. Additional case-specific modifications:
- a. There was only one medium facility with 11 to 25 beds that had 100% of its resident living quarters be two-bedroom or three-bedroom apartments. In order to mask this case, we decided to recode the facility to have some of its resident living quarters also be studio or one-bedroom apartments.
 - b. Similarly, only a few small facilities with 4 to 10 beds had their resident living quarters equipped with a cooktop/hot plate or an oven. In order to mask these cases, they were coded as not having these amenities. There were no statistically significant differences in estimates as a result of this modification. However, this

may lead to non-statistically significant differences in estimates of these variables between the PUF and the NCHS' restricted, in-house file.

Variables with “other specify” responses

Respondents could select from a list of 12 reasons for refusing to admit and discharge residents. In addition, residents could specify other reasons that were not listed.

1. Some respondents indicated that the facility would refuse to admit someone for reasons other than the 12 reasons the interviewer asked about (ADOTH), and were asked to specify what the reason was (ADSPEC). Responses to ADSPEC were examined and responses were recoded to the 12 existing reasons if they matched them (ADEMER, ADCOG, ADBEH, ADSNC, ADMON, ADURINE, ADFECES, ADBOTH, ADLIFT, ADDRUG, ADEOLC).
2. Similar approach was taken and recoded to the existing 12 discharge reasons if other reasons that the respondent specified for DCSPEC matched the existing reasons (DCEMER, DCCOG, DCBEH, DCSNC, DCMON, DCURINE, DCFECES, DCBOTH, DCLIFT, DCDRUG, DCEOLC).

List of restricted variables

A complete list of the facility PUF variables that are masked or restricted is available at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010FacilityRestrictedvariables.pdf. Users wishing to access these restricted variables or link the facility PUF to the resident PUF or non-NCHS data files (e.g., Area Resource File) 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 also 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 facility 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 PUFSTRATA;	TOTCNT PUFPOPFAC;	WEIGHT FACFNWT;

Table 1b. Computations using STATA

Design description in STATA
svyset facid [pweight=facfnwt], strata(pufstrata) fpc(pufpopfac) vce(linearized) singleunit(missing)

Table 1c. Computations using SAS

PROC	STRATA	CLUSTER	WEIGHT
PROC SURVEY_ DATA = Y TOTAL = SECONDFILE;	STRATA PUFSTRATA;	CLUSTER FACID;	WEIGHT FACFNWT;

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 “PUFSTRATA” from the variable list as **Strata**.
 - Select “FACFNWT” 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 “PUFPOPFAC” for **Read values from variable**
5. Stage 1: Plan Summary
 - Select “No, do not add another stage now” for **Do you want to add stage 2?**

Log will be created as shown below:

```
CSPLAN ANALYSIS  
/PLAN FILE='<file name>'  
/PLANVARS ANALYSISWEIGHT=FACFNWT  
/PRINT PLAN  
/DESIGN STRATA=PUFSTRATA  
/ESTIMATOR TYPE=EQUAL_WOR  
/POPSIZE VARIABLE=PUFPOPFAC.
```

Obtaining the data

NSRCF data files are available 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.