2022 National Post-acute and Long-term Care Study

Residential Care Community Services User (Resident) Public Use Data File

Data Description and Usage

September 2024

Table of Contents

Please Read Carefully Before Using NCHS Public Use Survey Data

The National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), conducts statistical and epidemiological activities under the authority granted by the Public Health Service Act (42 U.S.C. § 242k). NCHS survey data are protected by Federal confidentiality laws including Section 308(d) Public Health Service Act [42 U.S.C. 242m(d)] and the Confidential Information Protection and Statistical Efficiency Act or CIPSEA [44 U.S.C. 3561-3583]. These confidentiality laws state the data collected by NCHS may be used only for statistical reporting and analysis. Any effort to determine the identity of individuals and establishments violates the assurances of confidentiality provided by federal law.

Terms and Conditions

NCHS does all it can to assure that the identity of individuals and establishments cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the dataset. Any intentional identification or disclosure of an individual 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 attempt to learn the identity of any person or establishment included in these data. Make no use of the identity of any person or establishment discovered inadvertently and inform the Director of NCHS, of any such discovery.
3. Not link this dataset with individually identifiable data from other NCHS or non-NCHS datasets.
4. Not engage in any efforts to assess disclosure methodologies applied to protect individuals and establishments or any research on methods of re-identification of individuals and establishments.

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

Data users are encouraged to report apparent errors in the RCC resident data or documentation files to the Data Analytics and Production Branch, Division of Health Care Statistics (ltcsbfeedback@cdc.gov).

Sanctions for Violating NCHS Data Use Agreement

Willfully disclosing any information that could identify a person or establishment in any manner to a person or agency not entitled to receive it, shall be guilty of a class E felony and imprisoned for not more than 5 years, or fined not more than $250,000, or both.

Obtaining the data

The RCC 2022 provider public use file is available free of charge to users from the NPALS website (https://www.cdc.gov/nchs/npals/questionnaires/index.html).

We appreciate users inform the Data Analytics and Production Branch of any publications or presentations produced based on the 2022 NPALS data and cite relevant NPALS documentations/data products in their work when appropriate.

Introduction

This document describes the residential care community (RCC) services user (or resident) public use data file and the processes involved in creating it. NCHS recommends that data users read this document prior to working with the data.

The National Study of Long-Term Care Providers (NSLTCP) was renamed the National Post-acute and Long-term Care Study (NPALS) in January 2020.

Data files

The 2022 NPALS RCC public use data are distributed in two files: (1) a provider-level data file and (2) a services user (resident)-level data file. This document refers to the services user public use data file. The file contains one record for each sampled resident (residents were not interviewed; interviews were conducted with the directors or caregivers or knowledgeable RCC staff). This file covers data on resident characteristics, resident use of services, resident health status, and resident cognitive and physical functioning. The provider identifier (RCCID) in the provider file is different from the provider identifier (SU_FACID) in the services user file and cannot be used to link the provider and services user files. This file has 549 records and 83 variables, each with a primary identifier (SUID) and sorted in order by SUID and SU_FACID.

The data file is provided in ASCII, with fixed-length records, SAS, STATA, and R formats. Public use data files can be downloaded from the NPALS website as separate files. The individual files for separate download are:

Documentation files	
Survey methodology documentation	https://www.cdc.gov/nchs/data/npals/2022-Survey-Methodology-Document.pdf
Data dictionary	https://www.cdc.gov/nchs/data/npals/2022-RCC-Services-User-PUF-codebook.pdf
Services User Questionnaire	https://www.cdc.gov/nchs/data/npals/2022-RCC-Services-User-Questionnaire.pdf
This document (ReadMe file)	https://www.cdc.gov/nchs/data/npals/2022-RCC-Services-User-PUF-ReadMe.pdf

Documentation

This RCC Readme file is part of the documentation package accompanying the release of the 2022 RCC services user (resident) public use file. The package also includes a data dictionary or codebook, and the services user or resident questionnaire. A broader NPALS survey methodology document is also included in the public use data file release package.

Brief description of survey and outcomes

The RCC resident survey was conducted between September 2022 and March 2023. To be eligible for the study an RCC had to be licensed, registered, listed, certified, or otherwise regulated by the state; had four or more licensed, registered, or certified beds; provided room and board with at least two meals a day, around-the-clock on-site supervision, and help with personal care, such as bathing and dressing or health related services such as medication management. RCCs had to serve a predominantly adult population. RCCs licensed to exclusively serve the mentally ill or the intellectually disabled/developmentally disabled populations were excluded from NPALS. Services user data were collected via computer-assisted telephone interviews (CATI) and RCC provider level data were collected by mail, web, and CATI.

A two-stage sampling design was used for the 2022 NPALS RCC component. At the first stage, a stratified sample of RCCs were selected and at the second sampling stage, samples of two residents were selected from each eligible participating sample provider. From a frame of 46,049 RCCs, 2,088 RCCs were selected for the survey. Of the 2,088 RCCs in the sample, 1,326 RCCs could not be contacted and, therefore, the eligibility status of these RCCs was unknown. Using the eligibility rate of 90% derived from RCCs that completed the screening questions, a proportion of RCCs of unknown eligibility was estimated or “presumed” to be eligible. This estimated number along with the total number of eligible RCCs were used to estimate the total number of eligible RCCs. Of the 1,885 eligible and presumed eligible RCCs, 688 of them completed the provider questionnaire, for a weighted response rate (for differential probabilities of selection) of 34.0% calculated using AAPOR’s Response Rate 4 approach (The American Association for Public Opinion Research, 2023), resulting in an estimated national total of 32,231 RCCs. The SU module was completed for 549 residents for a response rate of 40%, resulting in an estimated 1,016,424 residents. To account for the RCCs of unknown eligibility, the weights of the RCCs with known eligibility were adjusted upward based on the proportion of communities that were actually known to be eligible. Adjustments were also made to account for non-response.

Two residents were selected from eligible RCCs that agreed to participate. A questionnaire was completed via CATI for a total of 549 residents from 276 RCCs. About 40% (276 out of 688) of RCCs completed both provider and at least one SU module and the rest of the RCCs either completed the provider module or the SU module. Therefore, not all RCCs in the provider public use file are represented in the resident public use file and vice-versa. Some RCCs only completed one resident questionnaire. During the weighting process, these residents were moved to a similar RCC that had user records in the file.

Data dictionary

The 2022 RCC resident data dictionary or codebook for the public use file is provided as a single file containing all variables in the resident public use file. Each variable in the public use file has its own codebook entry.

If a question or a series of questions in the survey were legitimately skipped for selected respondents, then the skipped responses were coded as "-1= Inapplicable" in the data dictionary. The questionnaire skip pattern is specified in the data dictionary beside the question text and code categories. Data users are advised to consult the questionnaire to better understand the questionnaire skip patterns. Missing responses were coded as "-9=Not ascertained." The data dictionary or codebook is available on NPALS website and available upon request (ltcsbfeedback@cdc.gov).

Resident questionnaire

The Resident Questionnaire is included in the data release package and available at: https://www.cdc.gov/nchs/npals/questionnaires/index.html.

The questionnaire includes all the questions asked in the resident module. There may be some differences in how questions were asked in the questionnaire and how they were coded in the public use file. 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 public use file may have fewer response categories (response categories collapsed) than the number of categories indicated in the questionnaire. Restricted versions of the variables and data are available to users through the NCHS Research Data Center (http://www.cdc.gov/rdc/index.htm).

Data processing activities to create the public use file

The raw data received from the field were reviewed and edited prior to creating and releasing the public use file. Data were reviewed for accuracy, logic, consistency, and completeness. Additionally, extensive

disclosure risk review was conducted to prevent the identity of any facilities that 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. The following methods were employed on the restricted in-house file to create the public use file.

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 CATI, or the interview broke off before administering the entire questionnaire. In the data file, item-nonresponse is coded as -9 when a respondent did not provide an answer. NCHS handled item nonresponse for age (4 cases), sex (1 case), and race and ethnicity (10 cases) by imputing.

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 file. In making these modifications, NCHS staff tried to maintain a balance between the need for data confidentiality and the needs of data users.

1. Direct identifiers such as names, addresses, and geographic information (region, state, metropolitan statistical area) are not included in the public use file, and some other variables were not included in the public use file. Full list of restricted variables can be obtained by request (ltcsbfeedback@cdc.gov) or through the NCHS research data center (http://www.cdc.gov/rdc/index.htm).
2. Provider characteristics in the provider public use file are not included in the services user public use file (e.g. Ownership, Chain).

Modified variables

1. Some categorical variables were collapsed into fewer response categories. These included RACEETHR**,** MOVEFROMR, MEDICAIDR, MEMORYR, SIGHTR, COMUNICATR, INCONTR, FALLINJR, and SYMPTOMSR.
2. Length of stay in facility: Instead of including the original 5 categories when a sampled resident moved into a facility, a 3-category collapsed length of stay variable (LOSrc) is provided to reduce disclosure risk.

3. Race of resident: American Indians and Alaska Natives (AIAN), Hawaiian or other Pacific Islanders (NHOPI), Asian (ASIAN), Black (BLACK), other race (OTHERRACE) and multiple race (MULTIRACE) residents were collapsed into "Other" race/ethnicity category so that the race/ethnicity variable (RACEETHR) has only two categories: White non-Hispanic and Other.
4. Charges: The continuous variable was dichotomized (CHARGESR), and a two-category variable is provided.
5. Number of falls: The continuous variable was dichotomized (FALLNUMR) and a two-category variable is provided.

Top/bottom coded variables

For the variable AGE, 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 were included in the top and bottom categories of the recoded age variable (AGER).

Reliability of estimates

Estimates published by NCHS must meet reliability criteria published in two NCHS reports: "National Center for Health Statistics Data Presentation Standards for Proportions" is available from https://ww.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf and "National Center for Health Statistics Data Presentation Standards for Rates and Counts" is available from https://ww.cdc.gov/nchs/data/series/sr_02/sr02-200.pdf. Estimates not meeting NCHS standards are not presented or are flagged based on the procedure specified in these guidelines. Users of the NPALS public use files are encouraged to assess the reliability of estimates derived from their analyses, though they are not required to use NCHS guidelines.

Analyses and Weighting of NPALS public use files

The data collected in the 2022 NPALS were obtained through a complex, multistage sample design that involved stratification and clustering. The final weights provided for analytic purposes have been adjusted in several ways to yield valid national estimates for residential care communities in the U.S. Users are reminded that the use of standard statistical procedures 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 NPALS provider public use file or NPALS SU public use file. The clustering protocols that are used in the multistage selection of the NPALS sample require

additional analytic procedure consideration as described below. Users who apply SRS techniques to the data 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 code for SUDAAN, SAS, Stata, and R software packages are provided for illustrative purposes (Tables 1a-d). However, the appropriate application of these procedures is the ultimate responsibility of users. NCHS strongly recommends that NPALS data be 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 RCC provider public use file includes design variables that designate each record's stratum marker and the first-stage unit (or cluster) to which the record belongs. The design variables in the public use file are masked and not the same as the design variables in the restricted data files.

Table 1a. Computations using SUDAAN

PROC statement	NEST statement	TOTCNT statement	WEIGHT statement
PROC x FILE = y DESIGN = WOR;	NEST PUFSTRATA SU_FACID;	TOTCNT PUFPOPFAC POPSU;	WEIGHT SUWT;

Table 1b. Computations using STATA

Design description in STATA
svyset su_facid, strata(pufstrata) fpc(pufpopfac) vce(linearized) singleunit(missing)||suid, fpc (popsu) weight (suwt)

Table1c. Computations using SAS

PROC	STRATA	CLUSTER	WEIGHT
PROC SURVEY_ DATA = Y TOTAL = SECONDFILE;	STRATA PUFSTRATA;	CLUSTER SU_FACID;	WEIGHT SUWT;

Table 1d. Computations using R

Design description in R (with R package 'survey')

```
# Create design object (use options for handling lonely PSUs)
options(survey.lonely.psu = "adjust")  #to address lonely PSUs or see package 'survey' usage documentation for alternatives.

design_object <- svydesign(id=~SU_FACID+SUID, strata=~PUFSTRATA, weights = ~SUWT, data = SUDATA, fpc=~PUFPOPFAC+~POPSU, nest = TRUE)

#Use svymean()to obtain proportions and standard errors of categorical variables:

svymean(~VARIABLE, design_object)
```

Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)

The table below provides a Preferred Reporting Items for Complex Survey Analysis (PRICSSA) document (Seidenberg, Moser, & West 2023) for users of the 2022 NPALS residential care community (RCC) services user (resident) public use data file. This information may be helpful to users when analyzing the 2022 NPALS survey data.

Table 2. Preferred Reporting Items for Complex Sample Survey Analysis

Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)	Description
Name of survey	National Post-acute and Long-term Care Study Residential Care Community Component
Data collection mode	Mail or web with telephone follow-up
Target population	Residential care community residents (through their directors or knowledgeable staff) in the United States
Populations excluded	RCCs licensed to exclusively serve severely mentally ill or intellectually disabled/developmentally disabled populations
Variance and standard error estimation	Taylor Series Linearization
Sample design	Stratified random sample
Weight	SUWT
Design variable: Stratum	PUFSTRATA
Design variable: population correction factor	PUFPOPFAC
Design variable: Facility ID in the user file for estimation	SU_FACID
Design variable: Total number of services users in each facility ID	POPSU
Presentation standards	Proportions or percentages:https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf Rates and counts: https://www.cdc.gov/nchs/data/series/sr_02/sr02-200.pdf
Unweighted total sample size	549 RCCs
Weighted total sample size	1,016,424 residents
Response rate (weighted)	38.0%
Location of example code	See Table 1 (a-d) above for approaches in various statistical analysis programs.

Suggested citation

Residential care community services user (resident) public use README file (this document):

National Center for Health Statistics. Division of Health Care Statistics. *2022 National Post-acute and Long-term Care Study (NPALS). Residential care community services user (resident) public use data file description and usage*, September 2024. Hyattsville, Maryland.

Residential Care Community services user data file:

National Center for Health Statistics. Division of Health Care Statistics. *2020 National Post-acute and Long-term Care Study (NPALS). Residential care community services user (resident) public use data file*, September 2024. Hyattsville, Maryland.

Contact Information

For questions, suggestions, or comments concerning NPALS data, please contact the Data Analytics and Production Branch at:

Data Analytics and Production Branch, National Center for Health Statistics,
3311 Toledo Road, Hyattsville, MD 20782
E-mail: ltcsbfeedback@cdc.gov
Phone: 301-458-4747

Reference

1) The American Association for Public Opinion Research. 2023 Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys. 10th edition. AAPOR.
2) Seidenberg AB, Moser RP, West BT. Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA). *Journal of Survey Statistics and methodology* 2023; 11(4):743-757