2022 National Post-acute and Long-term Care Study

Residential Care Community Provider 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 stated requirements.

Data users are encouraged to report apparent errors in the RCC provider data or documentation files to Division of Health Care Statistics, Data Analytics and Production Branch. (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 data file is available free of charge from the NPALS website (https://www.cdc.gov/nchs/npals/questionnaires/index.html).

We also appreciate users inform the Division of Health Care Statistics, Data Analytics and Production Branch of publications or presentations based on the 2022 NPALS data and cite relevant NPALS documentations/data products in their work when appropriate.

Introduction

The purpose of this document is to describe the 2022 National Post-acute and Long-term Care Study (NPALS) residential care community (RCC) provider public use data file and some of the processes involved in creating the RCC provider public use data files. We recommend that data users read this document prior to working with the public use data file.

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 data files: (1) a provider level and (2) a services user (resident) level data. This document describes the RCC provider public use file. The provider file contains one record for each sampled and eligible RCC that completed a provider questionnaire. The provider public use file covers characteristics about RCCs, services they provided, types of staff employed, and some aggregate resident characteristics. The provider file has 688 records and 100 variables. Each record contains a primary identifier (RCCID). The records in the provider file are sorted in order by the primary identifier. Note that the provider and services user public use data files cannot be linked using the identifiers provided in the public use data files.

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

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

Documentation

This RCC Readme file is part of the documentation package accompanying the release of the 2022 RCC provider public use file. The package also includes the broader NPALS survey methodology document, a data dictionary or codebook, and the provider questionnaire.

Brief description of survey

The RCC 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. Eligibility was determined by a series of questions in the provider questionnaire; RCCs that were screened out did not complete the rest of the questionnaire. Data were collected by mail, web, and computer-assisted telephone interviews (CATI). The intended respondents were directors/owners/administrators or otherwise knowledgeable staff of the RCC.

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% using AAPOR's Response Rate 4 calculation methods (The American Association for Public Opinion Research, 2023), resulting in an estimated national total of 32,231 RCCs.

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.

Data dictionary

The 2022 RCC provider data dictionary (i.e. codebook) for the public use file is provided as a single file containing information from each section of the provider questionnaire: A) Background Information; B) Services Offered; C) Resident Profile; D) Staff Profile; and E) Information on COVID-19. Each variable in the public use data file has its own codebook entry.

If a question or a series of questions in the survey were legitimately skipped for selected respondents, responses were coded as "-1= Inapplicable" in the data dictionary. The question 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 question skip patterns. Missing responses were coded as "-9=Not ascertained." The data dictionary is posted to the NPALS website as part of the 2022 public use file release.

Provider questionnaire

The Provider Questionnaire is available at: https://www.cdc.gov/nchs/npals/questionnaires/index.html

The questionnaire includes all the questions asked in the provider module. There may be some differences in how questions were asked in the questionnaire and how they are 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 provide percentages for some variables (e.g., percent of residents in various age categories, 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 residents younger than 65, residents age 65 to 74, number of residents with some or all of their long-term care services paid by Medicaid). The variables included in the list of restricted variables 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 or their residents 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. The following methods were employed to create the public use data file.

Consistency checks

1. To ensure internal consistency of the data, for some questions, edit checks were programmed into the web questionnaire and CATI system and 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 web questionnaire and CATI system prompted respondents and interviewers, respectively, to verify if the total number of residents provided by the respondent was accurate when it was not within ±10% range of the total number of residents reported in an earlier question.
2. In most cases, the same skip logic that was applied to the web questionnaire was used to edit the data file when the skip instruction was not followed by a respondent. For instance, if the respondent indicated that the RCC only served adults with Alzheimer disease or other dementias (Question 10) but provided responses to or left Questions 11 and 12 blank, then Questions 11 and 12 were recoded as "-1=INAPPLICABLE". However, if the response to Question 10 was missing and Questions 11 and 12 had a response, then Question 10 was recoded to '2=No'.
3. The variables for the age and race/ethnicity distribution of residents were edited if the values did not add to the total number of residents (Question 4). For example, when values for the age categories of an RCC (Question 28) did not add up to the total number of residents (Question 4), values were adjusted to sum to the total number of residents based on the proportion of values reported for different age categories for the case. The public use file does not include a variable indicating total number of residents. However, the public use file includes the age and collapsed race/ethnicity variables converted into percentages using the total number of residents as the denominator.
4. Ownership and Chain status (Question 1 OWNERSHIPrc and Question 14 CHAIN): When a case was missing a value for ownership or chain status in the survey data file but had a value in the sampling frame, then the missing value on the survey data file was recoded to the value of the variable in the sampling frame.

Changes in data because of respondent comments

The NPALS Web and CATI provider questionnaires allowed respondents to enter comments by clicking an icon provided for each question on each screen. For hard-copy questionnaires, keyers entered any notes respondents wrote in the margins or in response boxes as they keyed the data. These comments were compiled and reviewed. The original response to the survey question was changed if it was determined that the comment changed the substance of the recorded answer.

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. There were other variables that were not included in the public use file. Full list of restricted variables is available through the RDC (http://www.cdc.gov/rdc/index.htm) or by request (ltcsbfeedback@cdc.gov).
2. Several variables have been modified to minimize disclosure risk.
    a. The total number of beds (Question 3) and the current number of residents (Question 4) in an RCC are not provided in the public use file but replaced by a 2-category occupancy rate variable (OCCU_CAT).
    b. Instead of providing only the different providers (physician, pharmacy, hospital, skilled nursing facility etc., and other long-term care provider) with which an RCC's electronic health records (EHR) system supported electronic health information exchange (Question 20a-e), two variables were derived to indicate any exchange with physician or pharmacy (ANYEX) and any exchange with hospital, or skilled nursing facility, or other long-term care provider (ANYIT).

Edited/derived variables

1) Hours per resident day, by employee staff type (i.e., RNHPPD, LPNHPPD, AIDEHPPD, SOCWHPPD, and ACTHPPD):
    i) Hours per resident day were derived from the number of full-time equivalents for each staff type and the current number of residents (Question 4). We identified outliers and

recoded as needed.

ii) The number of FTEs for a given employee staff type was then converted into hours by multiplying the FTEs by the average number of hours in a work week (based on a 35-hour work week) and dividing the total number of hours per staff type by the total number of residents and by the number of days in a work week (7 days). When HPPD variables had values greater than 24, these values were coded as 24.

2) Any employees by staff type (ANYRN_EMP, ANYLPN_EMP, ANYAIDE_EMP, ANYSOCW_EMP, ANYACT_EMP)

i) These binary variables were derived from the continuous FTE variables for employees (e.g., RNFT1 and RNPT1, LPNFT1 and LPNPT1, etc.…) and coded to indicate whether the RCCs had any employees of that staff type.

ii) These variables indicate the presence or absence ("1=YES", "2=NO") of any (full-time or part-time) employees.

Converting numbers to percentages

The provider public use file included several resident variables aggregated at the provider level (for example, AGE, etc.). Instead of providing the exact number of residents that fall within a category such as residents younger than 65 years, these variables were converted into percentages using the number of current residents (Question 4) as the denominator.

Item nonresponse

Item nonresponse is a source of missing data that occurred when a respondent did not know the answer to a question or refused to answer a question; or if the respondent submitted the questionnaire before all the questions were answered. The variables with the highest item-nonresponse were OMOTH (Question #27h: During the last 12 months, what did the (ombudsman) representative do for this residential care community?), followed by percent of residents in various age categories, race/ethnicity variables, and hours of training and continuing education variables, staffing and staff benefits variables, and COVID-19 related variables. Item nonresponse (weighted) was less than 10% for most variables.

List of restricted variables

Users wishing to access data with restricted variables or link the provider file to the services user file or non-NCHS data files (e.g., Area Resource File) need to contact the NCHS Research Data Center (http://www.cdc.gov/rdc/index.htm).

Reliability of estimates

Estimates published by NCHS must meet reliability criteria issued 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 Data 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 RCCs 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 the NPALS provider public use file. The clustering protocols that are used in the multistage selection of the NPALS sample require the use of other analytic procedures 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;	TOTCNT POPFAC;	WEIGHT FACWT;

Table 1b. Computations using STATA

Design description in STATA
svyset RCCID [pweight=FACWT], strata(PUFSTRATA) fpc(POPFAC) vce(linearized) singleunit(missing)

Table1c. Computations using SAS

PROC	STRATA	CLUSTER	WEIGHT
PROC SURVEY_ DATA = Y TOTAL = SECONDFILE;	STRATA PUFSTRATA;	CLUSTER RCCID;	WEIGHT FACWT;

Table 1d. Computations using R

Design description in R (with package 'survey')
Create design object: *design_object* <- svydesign(id=~**RCCID**, weights=~**FACWT**, strata=~**PUFSTRATA**, nest=TRUE, fpc=~**POPFAC**, data = RCCDATA) #Use svymean() function 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) provider public use data file. This information may be helpful to users when analyzing the 2022 NPALS survey data files.

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 communities (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	FACWT
Design variable: Stratum	PUFSTRATA
Design variable: population correction factor	POPFAC
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	688 RCCs
Weighted total sample size	32,231 RCCs
Response rate (weighted)	34.0%
Location of example code	See Table 1 (a-d) above for approaches in various statistical analysis programs

Suggested citation

Residential care community provider public use data 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 provider public use data file description and usage*, September 2024. Hyattsville, Maryland.

Residential Care Community provider public use 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 provider 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, NCHS,
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