2025 NATIONAL POST-ACUTE AND LONG-TERM CARE STUDY

RESIDENTIAL CARE COMMUNITY PROVIDER RESTRICTED DATA FILE DATA DESCRIPTION AND USAGE

Division of Health Care Statistics

National Center for Health Statistics

June 2026

Suggested Citation

Residential care community restricted data file data description and usage (this document):

National Center for Health Statistics. Division of Health Care Statistics. *2025 National Post-acute and Long-term Care Study (NPALS). Residential care community provider restricted data file description and usage*, June 2026. Hyattsville, Maryland.

Residential care community provider restricted data file:

National Center for Health Statistics. Division of Health Care Statistics. *2025 National Post-acute and Long-term Care Study (NPALS). Residential care community provider restricted data file*, June 2026. Hyattsville, Maryland.

Contact Information

Data users can find the latest information about NPALS on the website, at: https://www.cdc.gov/nchs/npals/index.html. For questions, suggestions, or comments concerning NPALS data, please contact the Data Analytics and Production Branch, Division of Health Care Statistics at:

Division of Health Care Statistics, National Center for Health Statistics

3311 Toledo Rd, Hyattsville, MD 20782

E-mail: ltcsbfeedback@cdc.gov

Phone: 301-458-4747

Contents

Please Read Carefully Before Using NCHS Restricted 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 [Pub. L. No. 115-435, 132 Stat. 5529 § 302]. 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.
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.

Report apparent errors in the data or documentation files to the Division of Health Care Statistics (DHCS) at 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 2025 NPALS restricted RCC provider data file can be accessed through NCHS' Research Data Center (RDC): https://www.cdc.gov/rdc/. In addition to following the RDC procedures for restricted data file access, the above terms and conditions also apply.

We appreciate users informing DHCS of any publications or presentations based on the 2025 NPALS data and cite relevant NPALS documentations/data products in their work when appropriate.

Residential Care Community Provider Restricted Data File Data

Description and Usage

In the following sections of the document, the introduction describes the objectives of the 2025 National Post-acute and Long-term Care Study (NPALS) and the README document for residential care communities (RCC) component of NPALS. Section two describes data files and related documentation. Section three describes data processing activities used to create the restricted data file, and section four briefly describes reliability of estimates.

Section 1 Introduction

The 2025 NPALS (renamed in 2020 from the National Study of Long-Term Care Providers, or NSLTCP) is designed to provide national and state representative statistical information about the supply and use of long-term care services providers in the United States. The main goals of NPALS are to: (1) estimate the supply of paid, regulated long-term care services providers, (2) estimate key policy-relevant characteristics and practices of these providers, (3) estimate the number of long-term care services users, (4) estimate key policy-relevant characteristics of long-term care services users, (5) produce national and state estimates, where feasible, within confidentiality and reliability standards, (6) compare across provider sectors, and (7) monitor trends over time. Data collection was conducted from November 2024 through July 2025. For convenience and because most of the data collection activities occurred in 2025, this wave is referred to as 2025 NPALS. This README document is for the RCC file and describes steps taken to prepare the restricted provider file.

Section 2 Data Files

The 2025 NPALS RCC restricted data are at the provider level only. This section describes the RCC provider data file, which has one record for each sampled and eligible RCC that completed a provider questionnaire. The data file contains characteristics about RCCs, services they provided, types of staff employed and contracted, and aggregate resident characteristics. The RCC provider data file contains 2,906 records and 264 variables. Each record in the RCC file has a primary identifier (CASEID). The 2025 NPALS RCC data are provided in SAS, Stata, and R data formats.

Section 2.1 Documentation

There are several types of documentation available for use with the data file. These include the survey methodology documentation that provides a brief overview of the survey, the data collection procedures, and the sampling design; the survey questionnaires; and this provider-specific data description and usage or README document available on the website: https://www.cdc.gov/nchs/npals/questionnaires/index.html.

Section 2.2 Brief Description of Survey

The survey on RCCs was conducted between November 2024 and July 2025. To be eligible for the study, an RCC (a) had to be licensed, registered, listed, certified, or otherwise regulated by the state within their specific licensure category, (b) must have had four or more licensed, registered, or certified residential care beds, (c) must have had at least one resident currently living in the residential care community, (d) must have provided room and board with at least

two meals a day, (e) must have provided around-the-clock on-site supervision, (f) must have provided help with personal care, such as bathing and dressing or health related services such as medication management, (g) had to serve a predominantly adult population, and (h) was not licensed to only serve adults with an intellectual or developmental disability, severe mental illness, or both. Data were collected through Mail and Web questionnaires. Computer-assisted telephone interviews (CATI) were not used in this wave but prompting calls were made towards the end of the survey period to selected RCCs to complete the Web or Mail questionnaire.

From a frame of 46,522 RCCs, 11,618 were selected for the survey. Of the 11,618 sampled RCCs, 3,793 responded, where 2,906 (76.6%) were eligible and 887 were ineligible as determined by the survey screening questions or were out of business. However, eligibility could not be determined for 7,825 of the sampled RCCs because they could not be contacted. Using the eligibility rate of 76.6%[1], a proportion of RCCs with unknown eligibility was estimated to be eligible. Thus, 5,994 RCCs of unknown eligibility were presumed to be eligible. The total number of eligible RCCs (responding eligible and presumed eligible) was estimated to be 8,900 (2,906+5,994). Of the 8,900 in-scope and presumed in-scope RCCs, 2,906 completed the provider questionnaire for a response rate of 32.6% (this is calculated by using AAPOR's Response Rate 4) [1]. To account for RCCs of unknown eligibility, the weights of the RCCs with known eligibility were adjusted upward based on the proportion of RCCs that were known to be eligible. Adjustments were also made to account for

[1] The eligibility rate is calculated by the number of known eligible RCCs divided by the total number of RCCs with known eligibility status. RCCs that were invalid or out of business and RCCs that screened out as ineligible were classified as known ineligibles.

nonresponse. In 2025, the estimated national number of RCCs is 31,416, serving an estimated 1,001,118 residents.

Table 1. Residential care communities sampling, response, and eligibility counts

Setting	Frame	Sample	Responded eligible	Responded ineligible	Total responded	Estimated eligible number of RCCs from RCCs of unknown eligiblility	Estimated Eligible + Responded Eligible
RCCs	46,522	11,618	2,906	887	3,793	5,994	8,900

Section 2.3 Data Dictionary

The 2025 RCC provider data dictionary (codebook) for the restricted data is provided as a single file. Each variable in the data file has its own entry in the data dictionary. Please note that some variables may not be available through RDC due to confidentiality and disclosure risk. A derived or transformed version of these variables may be available, where applicable.

If a question or a series of questions in the survey were legitimately skipped for selected respondents, then the skipped responses are 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 are coded as "-9=Not ascertained." The 2025 RCC Data Dictionary is available on the website: https://www.cdc.gov/nchs/npals/questionnaires/index.html.

Section 2.4 Provider Questionnaire

The RCC provider questionnaire is included in the data release package and available on the website: https://www.cdc.gov/nchs/media/pdfs/2024/11/NPALS-RCC-2024-quest-508.pdf. The questionnaire includes all the questions asked during the survey along with the skip patterns for selected questions. There may be some differences in how questions were asked in the survey and how they are coded in the restricted file. For example, the questionnaires use "mark all that apply" in questions that ask about different services that RCCs provide (Question 20a-k). Respondents indicated as many as two different ways that the RCC provided a given service, or if it is not provided. In the data file, for each service question, binary variables were included: three separate variables corresponding to three different ways that RCCs provide the service: 1) by paid RCC employees or by arranging for the service to be provided by outside service providers, 2) refers residents or family to outside service providers; and 3) does not provide, arrange, or refer for this service. In addition to these three binary variables, a derived variable with three mutually exclusive response categories is included in the data file for each service. These derived variables (with collapsed categories) indicate if the RCC provides the service: 1) by paid center employees/by arranging for the service to be provided by outside services providers; 2) only by referral; and 3) does not provide, arrange, or refer for the service.

The questionnaires have changed over the survey years. Therefore, for data users interested in multiple years of data, the previous questionnaires are available at:

https://www.cdc.gov/nchs/npals/questionnaires/index.html

Section 3 Data Processing Activities to Create the Restricted Data File

The raw data received from the field were reviewed and edited prior to releasing the restricted data file through NCHS' Research Data Center (RDC). Data were reviewed for accuracy, integrity, logical consistency, and completeness.

Section 3.1 Consistency Checks and Data Edits

1. To ensure internal consistency of the data, for some questions, edit checks were programmed into the Web questionnaire 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 prompted respondents to verify if the total number of male and female residents provided by the respondent was accurate when the sum of male and female residents did not add to the total number of residents reported in an earlier question.
2. The variables for sex, race-ethnicity, and age distribution of residents were edited if the values did not add to the total number of residents (Question 4). For example, when number of RCC residents by age categories (Question 23) did not add up to the total number of residents provided in Question 4, values in age categories were adjusted to sum to the total number of residents based on the proportion of values reported for different age categories for the case.
3. 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 9) but had indicated responses or left blank Questions 10 and 11, then Questions 10 and 11 were coded as "-1=Inapplicable". However, if the response to Question 9 was missing and Questions 10 and 11 had a response, then Question 9 was recoded to 'No'.

4. When a case was missing a response or value for ownership status (Question 1) in the survey data file but had a value for ownership in the sampling frame, then the missing value on the survey data file was recoded to the value of ownership on the sampling frame.

Section 3.2 Changes in Data Because of Respondent Comments

The NPALS Web questionnaire 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 was changed if it was determined that the comment changed the substance of the recorded answer.

Section 3.3 Edited/Derived Variables

1. Number of full-time and part-time employees, by employee staff type (Question 31a-e):

Number of full-time and part-time employees for a given staff type were edited to address cases with missing data. Instructions were provided in the questionnaire to enter "0" if the RCC had no employees for a given staff type. Yet, there were cases where respondents indicated the number of staff in the response box only when specific staff categories were applicable, while leaving inapplicable response boxes blank. Therefore, when editing full-

time/part-time staffing variables, blank responses were coded as "0" unless responses to all ten response boxes for all employee staff type were blank or missing (i.e., responses were blank for all employee staff types). Otherwise, if all response boxes were missing, then staffing categories were kept as missing (-9) because it could not be reasonably determined whether it is truly missing or the respondent meant "0" for each staff type. This coding scheme was similar to the scheme used in 2016, 2018, 2020, and 2022 but different from the coding scheme used in 2014.

2. Full-time equivalent employees (FTEs) – RNFTE1recode, LPNFTE1recode, AIDEFTE1recode, SOCWFTE1recode, ACTFTE1recode

Number of full-time and part-time employees for a given staff type were converted to number of full-time equivalents (FTEs) with an assumption that full-time is 1 FTE and part-time is 0.5 FTE. For example, an RCC with 3 full-time aide employees and 4 part-time aide employees would be considered to have 5 FTE aides.

Outliers for the FTE variables were defined as values that are 2 standard deviations above or below the size-specific mean for a given staff type, where size was defined as the number of current total residents categorized (1= 1-25 residents, 2=26-100 residents, 3=101 or more residents). Outliers were replaced by size-specific mean. When calculating the size-specific mean for a given staff type, cases were coded as missing if a) the number of FTE registered nurse employees/contract staff was greater than 999, b) the number of FTE licensed practical/vocational nurse employees was greater than 999, c) the number of FTE personal care aide employees was greater than 999, d) the number of FTE social work employees was greater than 99, and e) the number of FTE activities employees was greater than 99.

3. Hours per resident per day, by employee staff type – RNHPPD, LPNHPPD, AIDEHPPD, SOCWHPPD, and ACTHPPD
    a. Hours per resident per day were derived from the number of full-time equivalents for each type of staff and the current number of residents (Question 4).
    b. The number of FTEs for a given employee staff type was converted into hours by multiplying the FTEs by the average number of hours in a work week (based on a 35-hour work week). Then the total number of hours per staff type was divided by the total number of residents and by the number of days in a week (7 days for RCCs). When variables had values greater than 24 hours, these values were capped at 24.
4. Any employees, by staff type – ANYRN_EMP, ANYLPN_EMP, ANYAIDE_EMP, ANYSOCW_EMP, and ANYACT_EMP
    a. These variables were derived from the FTE variables for employees (e.g., RNFTE1recode was used to derive ANYRN_EMP) indicating whether the RCC had any RNs who are employees.
5. Electronic health records systems to exchange information with other providers

Having a computerized system that supports electronic health information exchange with any of the following providers or settings (ANYEX) was derived from the types of exchanges listed in Question 17a-e: physicians (ITMDrc), pharmacies (ITPHARMrc), hospitals (ITHOSPrc), nursing homes (ITNHrc), or other long-term care establishments (ITLTCOTHrc). The ANYEX variable was derived to indicate whether the RCC's electronic health records system supported electronic information exchange with any of the five entities in Question 17.

In addition to the above recoded or derived variables, several other variables were also

recoded. All recoded variables end with 'rc' to indicate the original variable was recoded (for instance, ITMDrc, ITHOSPrc etc.). In some cases, variables were derived from several other variables and the derived variables do not have 'rc' suffix. For instance, SERVHOSPR, SERVSOCWR, SERVMHR (and all other service variables) were created from the respective individual binary variables. As an example, SERVHOSP1, SERVHOSP3, and SERVHOSP4 were used to create SERVHOSPR as described in previous sections. Consult the data dictionary and other documentation for additional information on recoded or derived variables.

Section 3.4 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 answering all the questions. In the data file, item nonresponse or missing is coded as "-9= Not ascertained". Weighted nonresponse ranged from 0.9% for the ownership (OWNERSHPrc) variable (Question 1) to 20.3% for the asthma (DXASTHrc) variable (Question 27c).

Section 3.5 Imputed Data

Missing values for demographic variables age (Question 23a-d), sex (Question 24a-b), race-ethnicity (Question 25a-j), and two activities of daily living (ADL) variables – eating (26b) and bathing(26d) – were imputed.

After the weights were finalized, multiple imputations were created using the Cox-Iannacchione Weighted Sequential Hot Deck (WSHD) procedure in SUDAAN. For the WSHD

procedure in SUDAAN the variables used in the imputation procedure, referred to as the imputation class variables, must be specified. Within the cross of the imputation class variables, all responding and non-responding records for a given variable were identified. The responding records were potential donors for non-responding (missing) records. In other words, respondents were selected sequentially from within the cross of the imputation class variables and became donors for missing records within that same cross of variables.

For all demographic variables and two ADL variables (eating and bathing) with missing cases, class variables specified for the imputation procedure included: state, ownership type (OWNERSHP), chain affiliation (CHAIN), RCC authorized to participate in Medicaid (MEDICAID), metropolitan statistical area status (MSA), and RCC size (number of RCC beds, categorized). Cases with missing data were recoded as the mean of five imputed values for that specific case and cases with no missing data kept the value that respondents reported. The imputed race-ethnicity, age, sex, and ADL (eating and bathing) variables have 'rc' suffixed to the original variable name (e.g., FEMALErc, HISPANICrc, AG64LESSrc). Note that the rest of the ADL variables may have 'rc' suffixed to the original name because of other recoding done but were not imputed.

A flagging variable is also included to indicate number of cases imputed for the race-ethnicity, age, sex, and two ADL variables (e.g., AG64LESS_IMPFLG, FEMALE_IMPFLG). The percentage of imputed records ranged from 7.19% (209 missing responses) for the sex variables (Question 24a-b) to 15.2% (442 missing responses) for the 'Other race category' variable (Question 25i).

Section 4 Reliability of Estimates

Estimates published by NCHS must meet reliability criteria. Proportion estimates that do not meet the reliability criteria are not presented or are flagged based on the procedure specified in "National Center for Health Statistics Data Presentation Standards for Proportions," available on the website: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf. For totals and rates, reliability is determined by "National Center for Health Statistics Data Presentation Standards for Rates and Counts," available on the website: https://www.cdc.gov/nchs/data/series/sr_02/sr02-200.pdf. For example, for rates with sampling variability, if nominal or effective sample size in the numerator or denominator is less than 10, the estimate is suppressed. Similarly, if an estimated rate has a relative confidence width greater than 160%, the estimate is suppressed.

The RCC data collected in the 2025 NPALS were obtained through a complex survey design that involved stratification and clustering and a census of all RCCs in some states. 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 the NPALS RCC data were 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 file. The clustering protocols that are used in the multistage selection of the NPALS sample require other analytic considerations, as described below. Users who apply SRS techniques to the 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 rate.

In this document, examples of SUDAAN, Stata, SAS, and R specifications of analysis codes are provided for illustrative purposes (Tables 2a-2d). However, the appropriate application of these procedures and choice of analysis program are 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 file includes design variables that designate each record's stratum marker and the first-stage unit (or cluster) to which the record belongs. Examples for using these design variables with SUDAAN, Stata, SAS, and R survey procedures are provided below.

Table 2a. Computations using SUDAAN

PROC statement	NEST statement	TOTCNT statement	WEIGHT statement
PROC x FILE = Y DESIGN = WOR;	NEST STRATA;	TOTCNT POPFAC;	WEIGHT FACWT;

Table 2b. Computations using STATA

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

Table2c. Computations using SAS

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

Table 2d. Computations using R

Create design object:
design_object <- svydesign(id=~CASEID, weights=~FACWT, strata=~STRATA, nest=TRUE, fpc=~POPFAC, data = Y)
Package: “survey” svymean(~VARIABLE, design)
Package “surveytable” set_survey(design, mode = "NCHS")

Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)

The table below provides Preferred Reporting Items for Complex Survey Analysis (PRICSSA) [2] for users of the 2025 NPALS Residential Care Community (RCC) provider data. This information may be helpful to users when analyzing the 2025 NPALS survey data.

Table 3. 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 Provider File
Data collection mode	Mail or Web
Target population	Residential Care Communities 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 and census in some states
Weight	FACWT
Design variable: Stratum	STRATA
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	2,906 RCCs
Weighted total eligible number of RCCs	31,416
Weighted total current residents	1,001,118 residents
Response rate (weighted)	32.6%
Location of example code	See Table 2 (a-d) above for approaches in various statistical software.

Section 5 References

1. American Association for Public Opinion Research. 2023 Standard Definitions: Final Dispositions 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). J Surv Stat Methodol. 2023 Apr 26; 743-757.