2020 National Post-acute and Long-term Care Study

(Formerly called the National Study of Long-Term Care Providers)

Residential Care Community (RCC) Restricted Data File

July 2022

Data Description and Usage

Table of Contents

Please Read Carefully Before Working with the Data File

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.

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.

****

This document describes the data and the processes involved in creating the residential care communities (RCCs) provider restricted data file. 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 2020 NPALS RCC restricted data are at the provider-level. This document describes the RCC provider data file. The provider data file has one record for each sampled and eligible RCC that completed a provider questionnaire and contains characteristics about RCCs, services they provided, types of staff employed, and aggregate resident characteristics. The RCC provider data file contains 4,312 records and 242 variables. Each record in the RCC file has a primary identifier (CASEID) and records are sorted in the order of the primary identifier.

The 2020 NPALS RCC data are provided in SAS and STATA data formats.

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; this provider-specific data description and usage or readme document. A data dictionary or codebook listing the questions and response categories (without the unweighted frequencies and weighted estimates) will be available upon request.

Brief description of survey

The survey on RCCs was conducted between November 2020 and July 2021. 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; and (g) had to serve a predominantly adult population; and (h) was not licensed to exclusively serve the mentally ill or the intellectually disabled/developmentally disabled populations. Data were collected by mail, web, and computer-assisted telephone interviews (CATI).

From a frame of 44,201 RCCs, 11,618 were randomly selected for the survey. Of the 11,618 sampled RCCs, eligibility could not be determined for 5,818. Among those for which eligibility could be determined (5,800), 4,366 (75%) were eligible and 1,434 (25%) were ineligible because they did not meet the survey definition of RCCs or were out of business. However, 5,818 (or 50%) RCCs could not be contacted; therefore, the final eligibility status of these RCCs was unknown. Using the eligibility rate of 75.3%,[1] a proportion of these RCCs of unknown eligibility was estimated to be eligible; therefore, 4,380 RCCs of unknown eligibility were assumed to be eligible. The total number of eligible RCCs was estimated to be 8,746 (4,366 + 4,380). Of the 8,746 in-scope and presumed in-scope RCCs, 4,312 completed the provider questionnaire, for a weighted response rate (for differential probabilities of selection) of 45% (this is calculated by using AAPOR's Response Rate 4); 54 eligible RCCs did not complete the provider questionnaire. To account for the RCCs of unknown eligibility, the weights of the RCCs with known eligibility were adjusted upward based on the proportion of RCCs that were actually known to be eligible. Adjustments were also made to account for non-response.

[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.

Differences in the number of residential care communities over time

The estimate of the number of residential care community providers varied across the survey waves beginning with the National Survey of Residential Care Facilities (NSRCF) in 2010, NSLTCP (2012-2018) and NPALS (2020). Estimated number of residential care communities across the waves were: 31,100 in 2010 (NSRCF), 22,185 in 2012, 30,245 in 2014, 28,858 in 2016, 31,422 in 2018, and 30,577 in 2020 (NPALS). These differences in estimated number of residential care communities are likely attributable to differences in eligibility rates. Description of possible reasons for differences in eligibility rates across the survey waves beginning with NSRCF and all NSLTCP waves particularly between the 2010 NSRCF and the 2012 NSLTCP as well as between the 2012 and 2014 NSLTCP waves is provided in a recent publication (Sengupta, Singh and Melekin, 2022) which can be accessed at https://www.cdc.gov/nchs/data/series/sr_02/sr02-192.pdf. Briefly, the main reasons for ineligibility were RCCs exclusively serving populations with mental retardation/developmental disabilities or with severe mental illness, and not providing 24-hour supervision among other minor modifications of screener questions as well as question instructions.

Data dictionary

The 2020 RCC provider data dictionary (codebook) for the restricted data is provided as a single file containing all five sections of information in 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 data 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 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=MISSING."

Provider Questionnaire

The Provider Questionnaire is included in the data release package and available at: https://www.cdc.gov/nchs/npals/2020-NPALS-RCC-Questionnaire-Community.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 residential care communities provide (Question 22a-h). Respondents indicated as many as five different ways that the residential care community provided a given service. In the data file, for each service, five binary variables were included: three separate variables corresponding to three different ways that residential care communities provide the service (i.e., by paid residential care community employees, by arranging for the service to be provided by outside service providers, by referring participants or family to outside service providers); one variable indicating whether the RCC *temporarily* does not provide, arrange, or refer for this service; and one variable indicating whether the RCC does not provide, arrange, or refer for this service. In addition to these five binary variables, a derived variable with three mutually exclusive response categories (collapsed categories) is included in the data file for each service. These derived variables indicate if the RCC provides the service: 1) by paid residential care community employees/by arranging for the service to be provided by outside services providers; 2) only by referral; or 3) temporarily does not provide, arrange, or refer for this service/does not provide, arrange, or refer for the service.

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 to NCHS' Research Data Center (RDC). Data were reviewed for accuracy, logic, consistency, and completeness.

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 systems prompted respondents and interviewers, respectively, 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 . 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 13) but had indicated responses or left blank Questions 14 and 15, then Questions 14 and 15 were coded as "-1=INAPPLICABLE". However, if the response to Question 13 was missing and Questions 14 and 15 had a response, then Question 13 was recoded to 'No'.

3. The variables for sex and age distribution of residents were edited if the values did not add to the total number of residents (Question 7). For example, when number of RCC residents by age categories (Question 26) did not add up to the total number of residents provided in Question 7, 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.

4. Ownership (Question 2 OWNERSHPrc): When a case was missing a response or value for ownership 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.

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 was changed if it was determined that the comment changed the substance of the recorded answer.

Edited/ Derived variables

1 . Number of full-time and part-time employees, by employee staff type (Question 34a-e):

   a . The number of full-time and part-time employees for a given staff type were edited to address the cases with missing data. Instructions were provided in the questionnaire to enter "0" if the residential care community 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. Thus, when editing full-time/part-time (FT/PT) variables, these were coded as missing "0" unless responses to all ten response boxes for all employee staff type were blank or missing (e.g., for employees, if the number of full-time RN employees, the number of part-time RN employees, the number of full-time LPN employees, the number of part-time LPN employees, the number of full-time aide employees, the number of part-time aide employees, the number of full-time social worker employees, the number of part-time social worker employees, the number of full-time activities staff employees, and the number of part-time activities staff employees). Otherwise, the missing (-9) were kept as missing (-9). This coding scheme was similar to the scheme used in 2016 and 2018, but different from the coding scheme used in 2014. When editing the 2014 data, missing FT/PT variables were coded as "0" unless responses to all four response boxes for a given staff type were blank or missing (e.g., the number of full-time RN employees, the number of part-time RN employees, the number of full-time RN contract staff, the number of part-time RN contract staff). Otherwise, the missing (-9) were kept as missing (-9). In the 2014 scheme, each staff type was grouped and included both employees and contract staff.

2 . Hours per resident day, by employee staff type (i.e., RNHPPD1, LPNHPPD1, AIDEHPPD1, SOCWHPPD1, and ACTHPPD1):

   a. Hours per resident day were derived from the number of full-time equivalents for each staff type and the current number of residents (Question 7). 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 residents served based on current residents (1= 1-25 residents; 2=26-100 residents; 3=101 or more residents). Outliers were replaced with size-specific mean. When calculating the size-

specific mean for a given staff type, cases were coded as missing: if the number of FTE registered nurse employees was greater than 999; if the number of FTE licensed practical/vocational nurse employees was greater than 999; if the number of FTE personal care aide employees was greater than 999; if the number of FTE social work employees was greater than 99; and if the number of FTE activities employees was greater than 99. 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) 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.

3. Any employees (ANYRN_EMP, ANYLPN_EMP, ANYAIDE_EMP, ANYSOCW_EMP, ANYACT_EMP), by staff type:
   a. These variables were derived from the FTE variables for employees (e.g., RNFTE1 was used to derive ANYRN_EMP) indicating whether the RCCs had any RNs who are employees.
4. Having a computerized system that supports electronic health information exchange with physicians, pharmacies, hospitals, nursing homes, or other long-term care establishments (ANYEX):
   a. This variable was derived from ITMDrc, ITPHARMrc, ITHOSPrc, ITNHrc, and ITLTCOTHrc (Question 18a-c).

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-SERVHOSP4 and SERVHOSPT were used to create SERVHOSPR as described in previous sections (under Provider Questionnaire heading above). Consult the codebook and other documentation for further information.

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 mostly related to the COVID-19 questions and included CVDPPEQ3N95, CVDPPEQ4N95, CVDHOSPRESrc, CVDDEATHRESrc, CVDDEATHEMP, CVDHOSPEMP and OMOTH ranging from 10-16.2% missing (weighted). For all other variables, item nonresponse (weighted) was less than 10%.

Imputed data

In the data file, item nonresponse is coded as "-9= Not ascertained." Missing values for race-ethnicity (Question 27a-i), sex (Question 25a-b), and age (Question 26a-d) variables 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 must be specified; they are referred to as the imputation class variables. 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 with missing, class variables specified for the imputation procedure include 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 [BEDSrc]). 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 as respondents reported. The imputed race-ethnicity, age, and sex variables have 'rc' suffixed to the original variable name (e.g., FEMALErc, HISPANICrc, AG64LESSrc). A flagging variable is also included to indicate number of cases imputed for the race-ethnicity, age, and sex variables (e.g., AG64LESS_IMPFLG, FEMALE_IMPFLG). Among 4,312 respondents, the percentage of imputed records ranged from 2% (116 missing responses) for the sex variables (Question 25a-b) to 10.3% (503 missing responses) for the 'Native Hawaiian

and Other Pacific Islander' or NHOPI and 'Other race category' (Question 27f and 27h respectively).

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. 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 from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf. For all estimates other than estimates of proportions in the tables, estimates are not presented if they are based on fewer than 30 cases in the sample data, in which case only an asterisk (*) appears. Estimates based on 30 or more cases include an asterisk if the relative standard error of the estimate exceeds 30%.

The data collected in the 2020 NPALS were 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 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.

In this document, examples of SUDAAN computer code are provided for illustrative purposes. Examples are provided also for the SAS and STATA software packages. 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 file 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 STRATA;	TOTCNT POPFAC;	WEIGHT FACWT;

Table 1b. Computations using STATA

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

Table1c. Computations using SAS

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

Obtaining the data

The 2020 RCC provider file can be accessed through NCHS' Research Data Center (RDC). In addition to following the RDC procedures for restricted data file access, there are a few conditions or restrictions for data use, and they are as follows:

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 RCC provider 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 2020 NPALS data and cite relevant NPALS documentations/data products in their work when appropriate.

Contact Information

To request a codebook or for questions, suggestions, or comments concerning NPALS 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.