2018 National Study of Long-Term Care Providers

Residential Care Community (RCC) Restricted Data File

March 2021

Data Description and Usage

Table of 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.

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 some of the processes involved in creating the residential care communities (RCC) provider restricted data file. We recommend 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. For the remainder of this document NPALS will be referred to as NSLTCP in order to correctly match the name of the study when the 2018 surveys were fielded.

Data Files

The 2018 National Study of Long-Term Care Providers (NSLTCP) RCC restricted data are distributed in two data files: (1) provider-level and (2) services user (resident)-level data. This document describes the RCC provider file. The provider file contains one record for each sampled and eligible RCC that completed a provider questionnaire. The provider file contains characteristics about RCCs, services they provided, types of staff employed, and aggregate resident characteristics. The provider file contains 503 records and over 150 variables. Each record contains a primary identifier (CASEID). The records in the provider file are sorted in the order of the primary identifier.

The data are provided in SAS and STATA formats.

Documentation

There are several types of documentation available for use with the data file. These include the survey methodology documentation that 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. A separate data description and usage is available for the adult day services center component of NSLTCP. A data dictionary or codebook listing the questions and response categories (without the unweighted frequencies and weighted estimates) will be made available to data users upon request.

Brief description of survey

The survey on RCCs was conducted between July 2018 and February 2019. To be eligible for the study an RCCs 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 NSLTCP. Data were collected by mail, web, and computer-assisted telephone interviews (CATI).

From a frame of 43,770 RCCs, 2,090 were randomly selected for the survey. Of the 2,090 sampled RCC, eligibility could not be determined for 977. Among those for which eligibility could be determined (1,113), 857 (77%) were eligible and 354 (23%) were ineligible because they did not meet the survey criteria or were out of business. However, 977 RCCs (33%) could not be contacted; therefore, the final eligibility status of these RCCs was unknown. Using the eligibility rate of 77%,[1] a proportion of these RCCs of unknown eligibility was estimated to be eligible; 752 RCCs of unknown eligibility were assumed as eligible. The total number of eligible RCCs was estimated as 1,609 (857 + 752). Of the 1,609 in-scope and presumed in-scope RCCs, 503 completed the provider questionnaire, for a weighted response rate (for differential probabilities of selection) of 30% (this is calculated by using AAPOR's Response Rate 4). 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 2018 RCC provider data dictionary (codebook) for the restricted data is provided as a single file containing all four sections of information in the provider questionnaire: A) Background

[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.

Information; B) Resident Profile; C) Services Offered; and D) Staff Profile. 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, responses of these respondents to the question were coded as "-1= INAPPLICABLE" in the data dictionary. The question skip pattern is specified in the data dictionary besides the question text and code categories. The data users are advised to consult the questionnaire to better understand the question skip patterns. 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/data/nsltcp/2018-NSLTCP-RCC-Questionnaire-Community.pdf

The data file 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 questionnaires, and how they are coded in the restricted file. For example, the questionnaires use "mark all that apply" questions to ask about different services that residential care communities provide (Question 26a-k). Respondents indicated as many as four different ways that the residential care community provided a given service. In the data file, for each service, four binary variables were included: four separate variables corresponding to four different ways that adult day services centers 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 center provides the service in any of these ways or does not provide the service. In addition to these four binary variables, a derived variable with three mutually exclusive response categories is included in the data file for each service. These derived variables indicate if the center 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) does not provide, arrange, or refer 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 the 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 system prompted respondents and interviewers, respectively, to verify if the total number of male and female residents provided by the respondent was accurate when it was not within ±10% range of the total number of residents reported earlier.

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's disease or other dementias (Question 7) but had indicated responses or left blank Questions 8 and 9, then Question 8 and 9 were coded as "-1—INAPPLICABLE". However, if the response to Question 7 was missing and Questions 8 and 9 had a response, then Question 7 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 17). For example, when values for the age breakdown of an RCC (Question 19) did not total to the total number of residents, 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.
4. Ownership (Question 3 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 NSLTCP 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, by employee staff type (Question 28a-f):
    a. Number of full-time and the number of part-time employees for a given staff type were edited to address the cases with missing data. Instruction was provided in the questionnaire to enter "0" if the center 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 missing as "0" unless responses to all ten response boxes for all employee staff type were blank or missing (e.g., for employees, 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, but different from the coding scheme used in 2014. When editing the FT/PT variables in 2014, they were coded missing 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 17). 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 average daily attendance (1= 1-25 residents; 2=26-100 residents; 3=101 or more residents). Outliers were coded as the 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 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, or hospitals (ANYEX)

   This variable was derived from ITMD, ITPHARM, and ITHOSP (Question 14a-c).

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 provision of dental services, the staff turnover variables followed by the age of residents. Item nonresponse (weighted) was less than 10% for all other variables.

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 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 2018 NSLTCP 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 that are based on the assumption that data are generated via simple random sampling (SRS) generally will produce incorrect estimates of variances and standard errors when used to analyze data from the NSLTCP

provider file. The clustering protocols that are used in the multistage selection of the NSLTCP sample require other analytic procedures, 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 NSLTCP 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	STRATA STRATA;	CLUSTER CASEID;	WEIGHT FACWT;

TOTAL = SECONDFILE;			

Obtaining the data

The RCC 2018 provider file can be accessed through the 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 2018 NSLTCP data cite relevant NSLTCP documentations/data products in their work when appropriate.

Contact Information

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