**2014 National Study of Long-Term Care Providers (NSLTCP)**

**Residential Care Communities Survey**

**Restricted Data File**


**July 2015**

**Data Description and Usage (Readme)**

Table of Contents

**Please Read Carefully Before Working with the Data File**

This document describes the data and some of the processes involved in creating the restricted file of the 2014 National Study of Long-Term Care Providers (NSLTCP)' survey of residential care communities.  We recommend that data users read this document prior to working with the data.


## Data File

The data file contains information on 5,035 residential care communities that met eligibility criteria for the study and completed a questionnaire in one of the three modes: a hardcopy mail questionnaire, a web questionnaire, and a computer-assisted telephone interview (CATI).  Each record in the file is associated with a primary identifier (CASEID) representing one eligible community that completed the survey.  Also included in the data file are 351 variables, including CASEID and design variables, on community characteristics (e.g., ownership type, chain affiliation, number of beds, Medicaid participation, dementia specific care units); number of employee and contract nursing, social work, and activities staff;  provision of services; demographic and health characteristics of residents in the community (e.g., race-ethnicity composition of residents, residents needing any assistance with different activities of daily living, residents' hospitalization and use of emergency department); use of electronic health records and health information exchange; derived variables; and imputed variables.  The records in the file are sorted in the order of the primary identifier.  The data are provided in SAS format.


## Documentation

There are several types of documentation available for use with the data file.  These include a data dictionary or codebook; the survey questionnaire; the survey methodology documentation that provides a brief overview of the survey, the data collection procedures, and the sampling design; and this provider-specific data description and usage or readme document.  A separate readme document on data description and usage is available for the adult day services center component of NSLTCP.


### *Brief description of survey*

The survey on residential care communities was conducted between June 2014 and January 2015.  The sampling frame was constructed from lists of licensed residential care communities obtained from the state licensing agencies in each of the 50 states and the District of Columbia. The 2014 NSLTCP used the same approach to creating the sampling frame (Wiener et al., 2010; available at: http://aspe.hhs.gov/daltcp/reports/2010/sfconst.pdf) and the same definition of residential care community as was used for the 2010 National Survey of Residential Care Facilities (NSRCF) (Moss et al., 2011); the PDF for the report is available at: http://www.cdc.gov/nchs/data/series/sr_01/sr01_054.pdf).

To be eligible for the study, a residential care community must have:

- Been licensed, registered, listed, certified, or otherwise regulated by the state to;
- Provide 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;
- Had four or more licensed, certified, or registered beds;
- Had at least one resident currently living in the community; and
- Served a predominantly adult population.

Excluded were residential care communities licensed to exclusively serve individuals with severe mental illness or intellectual disability/developmental disability. Nursing homes were also excluded.

For the residential care community component, the 2014 NSLTCP used a combination of probability sampling and taking a census. Probability samples were selected in the states that had sufficient numbers of residential care communities to enable state-level sample-based estimation. A state was sampled if it had enough communities to enable state-level estimation, i.e., if it had a sufficient number of communities to attain at least 81 completions after inflating the sample size for the estimated ineligibility and nonresponse. In states with an insufficient number of residential care communities on the sampling frame, NCHS took a census of communities. From 40,583 communities in the sampling frame, 11,618 residential care communities were sampled, stratified by state and facility bed size. A set of screener items was used to determine eligibility. Through the screening process it was determined that 128 (1.1%) residential care communities were invalid or out of business. However, 5,380 communities (46.3%) could not be contacted; therefore, the final eligibility status of these communities was unknown. Using the eligibility rate,[1] a proportion of these communities of unknown eligibility was estimated to be eligible. This estimated number along with the total number of eligible communities resulting from the screening process was used to estimate the total number of eligible residential care communities in the United States. The weights of the communities with known eligibility were adjusted upward based on the proportion of communities that were actually known to be eligible to account for the residential care communities with unknown eligibility status. Data were collected through three modes: self-administered hardcopy questionnaires, self-administered web questionnaires, and CATI conducted by interviewers. The questionnaire was completed for 5,035 communities, for a weighted response rate (for

---

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

differential probabilities of selection) of 49.6%.  Sample weights were adjusted to add up to the estimated number of eligible residential care communities (30,200).

*Differences in the number of residential care communities between 2010, 2012, and 2014*
The estimate of the number of residential care community providers varied between the 2010 NSRCF and the 2012 NSLTCP.  NCHS assessed these differences and concluded that they were largely related to eligibility differences between 2010 and 2012.  While both surveys used the same eligibility criteria, overall screener-based eligibility dropped from 81.0% in the 2010 NSRCF to 67.1%[2] in the 2012 NSLTCP.  The drop in screener-based eligibility rate was most obvious for providers with small bed size (4 to 10 beds): a decrease from 63.6% in 2010 to 45.8% in 2012.  Given that the 2012 NSLTCP (n=11,690) had a much larger sample than the NSRCF (n=3,605), and that small bed size providers make up the largest proportion of all residential care communities, the lower eligibility rate in 2012 compared to 2010 among small sized residential care communities had a large effect on the differences in the eligibility rate for the two surveys.

The probable reason for the eligibility differences between 2010 and 2012 was the different data collection modes used in 2010 (i.e., interviewer-administered CATI screener followed by in-person interview for eligible communities) and 2012 (i.e., primarily respondent self-administered screener and questionnaire completed by mail or web) and the resulting differences in how the respondents who self-administered the questionnaire interpreted the eligibility questions.  In the 2012 NSLTCP, the most common eligibility criteria that providers, particularly small sized residential care communities, did not meet were related to the provision of on-site, 24-hour supervision.  Some respondents using the self-administered modes (i.e., hardcopy questionnaire, web questionnaire) likely did not fully comprehend this question and might have screened themselves out of the study erroneously.[3]  Limited cognitive testing was conducted to assess these eligibility questions, and findings supported this hypothesis.  To address these differences and improve the clarity of the eligibility questions, NCHS revised the eligibility question asking whether the residential care community provided 24-hour supervision (Question 6 in the 2014 questionnaire)[4]. Results from the 2014 wave indicated that the overall eligibility rate increased to 80.7%; similar to the 2010 NSRCF.  Although the overall eligibility rate was similar the 2010 NSRCF rate, the rates for all bed size categories were slightly lower compared with the 2010 NSRCF (Table 1) and may be attributed to mode differences between 2010 and 2014.  The

---

[2] The screener-based eligibility rate was computed based on residential care communities that completed the screening questions. Eligibility rate= completed eligible / (completed eligible + completed ineligible).
[3] For more information, see 2013 overview report
   (http://www.cdc.gov/nchs/data/nsltcp/long_term_care_services_2013.pdf) or 2012 residential care community readme document
   (http://www.cdc.gov/nchs/data/nsltcp/NSLTCP_RCC_Readme_RDC_Release.pdf).
[4] Question 4 on the 2012 questionnaire.

estimated national number of residential care communities have ranged from 31,100 in 2010 to 22,200 in 2012, and 30,200 in 2014. The number of beds have estimated at 971,900 in 2010, 851,400 in 2012, and 1,000,014 in 2014.

Table 1. Percentage of eligible residential care communities, by bed size and survey year

| Eligible communities | 2014 National Study of Long-Term Care Providers | 2012 National Study of Long-Term Care Providers | 2010 National Survey of Residential Care Facilities |
|---|---|---|---|
| Overall | 80.7 | 67.1 | 81.0 |
| Bed size | | | |
| Small (4-10 beds) | 65.3 | 45.8 | 63.6 |
| Medium (11-25 beds) | 81.0 | 68.5 | 82.8 |
| Large (26-100 beds) | 91.7 | 82.4 | 94.5 |
| Extra large (over 100 beds) | 93.8 | 85.5 | 95.9 |

SOURCES:CDC/NCHS, National Study of Long-Term Care Providers, 2014,2012, and National Survey of Residential Care Facilities, 2010

### *Questionnaire*

The PDF for the residential care community questionnaire of the 2014 NSLTCP is available at:http://www.cdc.gov/nchs/data/nsltcp/2014_NSLTCP_Residential_Care_Communities_Questionnaire.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 questionnaire, and how they are coded in the restricted file. For example, the questionnaire uses "mark all that apply" questions to ask about different services that residential care communities provide (Question 15a – j). Respondents indicated as many as four different ways that the residential care community provided a given service. In the data file, for each service, five binary variables were included: four separate variables corresponding to four different ways that residential care communities provide the service (i.e., by paid residential care community employees, by arranging for and paying outside vendors, by arranging for outside vendors paid by others, by referral); one variable indicating whether the residential care community provides the service in any of these ways or not. In addition to these five 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 residential care community provides the service: 1) by paid residential care community employees/ arranging for and paying outside vendors/ arranging for outside vendors paid by others; 2) only by referral; or 3) does not provide the service.

*Data dictionary*

The data dictionary or codebook is provided as a single file containing five sections in the questionnaire: Study Eligibility; Background Information; Services Offered; Staff Profile; Resident Profile; and Record Keeping. Each variable in the restricted file has its own codebook entry.

The web and CATI versions of the NSLTCP survey of residential care communities used for data collection allowed respondents to answer or interviewers to only ask questions specific to the individual communities, skipping questions that did not apply. The skip instructions found in the hardcopy questionnaire were identified by bold, text in a different color (e.g., in Question 13, the instruction reads "If you answered Yes, skip to question"), which directed respondents to skip questions that did not apply. If a question or a series of questions in the survey were legitimately skipped by a particular respondent, then the response was coded as "-1= Legitimate skip" in the data dictionary. Skip patterns are specified in the data dictionary, in addition to the question text and code categories. When respondents refused to answer, did not know the answer, or did not answer the question because of a breakoff (i.e., did not complete the survey), their responses to the question(s) were coded in the data dictionary as "-9= Not ascertained." The data users are advised to consult the questionnaire before analyzing the data to better understand the question skip patterns.

## Data Processing Activities to Create the Restricted 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. 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 and CATI system was used to edit the data file when the skip instruction was not followed. For instance, if respondents indicated that the residential care community was not authorized

or set up to participate in Medicaid (Question 10), then their responses to the number of residents in the last 30 days who used Medicaid to pay for some or all of their services received at the residential community (Question 10a) were coded as "-1= Legitimate skip." However, the response to Question 10 was missing and Question 10a had a response, Question 10 was recoded as 'No' if the response to Question 10a was '0'; Question 10 was recoded as 'Yes' if the response to Question 10a was greater than 0.

3. The variables for race-ethnicity, sex, and age distribution of residents were edited if the values did not add to the total number of residents (TOTRES for Question 5). For example, when a case had missing data for a given race-ethnicity category, then the mean of five imputed values for that specific case was used to assess if values of the race-ethnicity categories summed to TOTRES. When values did not total to TOTRES, values were adjusted to sum to TOTRES based on the proportion of values reported for different race-ethnicity categories for the case.
   a. In addition to the original variables, edited variables for race-ethnicity, sex, and age distribution of residents are provided in the data file. The edited variables are indicated by adding "RC" as the suffix to the variable name (e.g., MALERC, FEMALERC for Question 19).
   b. Edited values for some cases are in decimals because of the following: recoding cases with missing data to take an average of five values imputed for that specific case or making proportional adjustments to individual categories when values did not total to TOTRES.

*Edited/ Derived variables*
1. Number of full-time equivalents by employee staff type (i.e., Question 17: RNFTE1, LPNFTE1, AIDEFTE1, SOCWFTE1, ACTFTE1) and contract staff type (RNFTE2, LPNFTE2, AIDEFTE2, SOCWFTE2, ACTFTE2)
   a. These variables were derived and provided in the restricted file. Number of full-time and the number of part-time employees and contract staff for a given staff type (separately for employees and contract staff) were converted into the number of full-time equivalents (FTEs) with an assumption that full-time is 1 FTE and part-time is 0.5 FTE. Instruction was provided in the questionnaire to enter "0" if the residential care community had no employees or contract staff 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 deriving FTE variables, we 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, we kept the missing (-9) as missing (-9).

    b.   Outliers for FTE variables were defined as values 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 (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 full-time equivalent (FTE) registered nurse employees/contract staff was greater than 999; if the number of FTE licensed practical/ vocational nurse employees/contract staff was greater than 999; if the number of FTE personal care aide employees/contract staff was greater than 999; if the number of FTE social work employees/contract staff was greater than 99; and if the number of FTE activities employees/contract staff was greater than 99.

2.  Hours per resident day, by employee staff type (i.e., RNHPPD1, LPNHPPD1, AIDEHPPD1, SOCWHPPD1, ACTHPPD1) and, by contract staff type (i.e., RNHPPD2, LPNHPPD2, AIDEHPPD2, SOCWHPPD2, ACTHPPD2)

    a.   Hours per resident day were derived from the number of full-time equivalents for each staff type and the number of current residents (TOTRES). The number of FTEs for a given employee staff type/ contract staff type was converted into hours by multiplying the FTEs by the 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 number of current residents in the community 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 staff employee (ANYRN_EMP, ANYLPN_EMP, ANYAIDE_EMP, ANYSOCW_EMP, ANYACT_EMP), any contract staff (ANYRN_CON, ANYLPN_CON, ANYAIDE_CON, ANYSOCW_CON, ANYACT_CON), and any employee or contract staff (ANYRN_EMPCON, ANYLPN_EMPCON, ANYAIDE_EMPCON, ANYSOCW_EMPCON, ANYACT_EMPCON), by staff type

    a.   These variable were derived from the FTE variables for employees and the FTE variables for contract staff (e.g., RNFTE1 to derive ANYRN_EMP; RNFTE2 to derive ANYRN_CON; and both RNFTE1 and RNFTE2 to derive ANYRN_EMPCON) indicating whether the residential care community had any employee staff RN, any contract staff RN, and any employee or contract staff RN.

4.  Having a computerized system that supports electronic health information exchange with physicians, pharmacies, or hospitals (ANYEX)

a. This variable was derived from ITMD, ITPHARM, and ITHOSP (Question 29).


5. Diagnosed conditions (DXALZ2, DXDEP2, DXSMI2, DXDD2, DXCADIO2, DXDIAB2), activities of daily living (TRANSHELP2, BATHHELP2, EATHELP2, DRESHELP2, BEDHELP2, WALKHELP2)
   a. Instruction was provided in the questionnaire to enter "0" if the residential care community had no residents with diagnosed conditions listed in Question 21 or activities of daily living (ADL) limitations listed in Question 22. Yet, there were cases where respondents indicated the number of residents with a given diagnosed condition/ ADL limitation in the response box only when specific diagnosed conditions/ ADL limitation categories were applicable, while leaving inapplicable response boxes blank. We coded missing as "0" unless responses to all response boxes for Question 21 or Question 22 were blank or missing. Otherwise, we kept the missing (-9) as missing (-9).


*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; the interviewer inadvertently skipped a question due to problems relating to CATI; or if the interview broke off before the entire questionnaire could be administered. The variables with the highest item nonresponse were the staffing variables: Question 17c.a AIDEFT1 (8.6%, weighted), Question 17c.b AIDEPT2 (17.6%, weighted), Question 16b STAFFFT (22.9%, weighted), and Question 16a STAFFPT (34.4%, weighted).


*Imputed data*

In the data file, item nonresponse is coded as "-9= Not ascertained." Missing values for race-ethnicity (Question 18), sex (Question 19), and age (Question 20) variables, DXALZ (Question 21a), DXDEP (Question 21d), EATHELP (Question22b), and BATHHELP (Question 22d) were imputed. In addition to the original variables, five sets of imputed variables are provided in the data file. Imputed variables are indicated by adding "imp" as the prefix and a numeral as the suffix to the variable name (e.g., impBATHHELP_1, impBATHHELP_5). A flagging variable is also included to indicate cases imputed for the variable (e.g., BATHHELP_FL). Among 5,035 respondents, the percentage of imputed records ranged from 5.6% (257 missing responses) for the categories of sex variable (MALE, FEMALE) to 14.2% (653 missing responses) for the variable indicating the number of residents diagnosed with depression (DXDEP).

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

      a.  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 variable names ending with "RC" (e.g., HISPANICRC, AG85UPRC, BATTHHELPRC, DXDEPRC), class variables specified for the imputation procedure include: state, number of beds, ownership type, chain affiliation, Medicaid participation, and metropolitan statistical area status.

## 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 the number of sampled records on which the estimate is based. The RSE is a measure of variability and is calculated by dividing the standard error of an estimate by the estimate itself. The result is then converted to a percentage by multiplying by 100. Guidelines used by NCHS authors to determine whether estimates should be presented in tables of NCHS published data reports include the following:

- If the estimate is based on 60 or more sampled cases and the RSE is less than 30%, the estimate is reported and is considered reliable.
- If the estimate is based on fewer than 30 sampled cases, the value of the estimate is not reported. This is usually indicated with an asterisk (*).
- All other reported estimates should not be assumed to be reliable. These include estimates with an RSE of 30% or more and estimates based on 30–59 cases, regardless of RSE.

The data collected in the 2014 NSLTCP residential care community (RCC) survey were obtained through a complex sample design that involves stratification and clustering. The final weights provided for analytic purposes have been adjusted in several ways to yield valid national and state estimates for RCCs in the United States. 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 2014 NSLTCP RCC survey. Users who apply SRS techniques to the 2014 NSLTCP RCC survey 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 and STATA computer code are provided for illustrative purposes. However, the appropriate application of these procedures is the ultimate responsibility of users. NCHS strongly recommends that users analyze the NSLTCP survey data under the direction of, or in consultation with a statistician who is knowledgeable in sampling methodologies and techniques for the analysis of complex survey data.

**Table 1a. Computations using SUDAAN**

| PROC statement | NEST statement | TOTCNT statement | WEIGHT statement |
|---|---|---|---|
| PROC x  FILE = y DESIGN = WOR; | NEST = FACSTRAT/ MISSUNIT; | TOTCNT= POPFAC; | WEIGHT= FACFNWT; |

**Table 1b. Computations using STATA**

| Design description in STATA |
|---|
|     svyset facid [pweight=facfnwt], strata(facstrat) fpc(popfac) vce(linearized) singleunit(missing) |

## Accessing the Restricted Data File

The 2014 NSLTCP RCC survey restricted data 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 only for statistical reporting and analysis.
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 data file or documentation to the Long-Term Care Statistics Branch (LTCSB).

We also request the user to inform LTCSB of any publications or presentations produced based on the 2014 NSLTCP RCC survey data, and cite relevant NSLTCP documentations/ data products in their work when appropriate.

## Contact Information

For questions, suggestions, or comments concerning the 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