2012 National Study of Long-Term Care Providers (NSLTCP)

Residential Care Communities Survey

Restricted Data File

October 2013

Data Description and Usage (Readme)

Table of Contents

Please Read Carefully Before Working with the Data File

****

The Public Health Service Act (Section 308 (d)) provides that the data collected by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), may be used only for the purpose of health statistical reporting and analysis. Any effort to determine the identity of any reported case is prohibited by this law. NCHS does all it can to assure that the identity of data subjects cannot be disclosed. However, the data released through the NCHS Research Data Center (RDC) include restricted variables, including geographic identifiers. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users will:

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.

By using these data, you signify your agreement to comply with the above-stated statutorily based requirements.

****

This document describes the data and some of the processes involved in creating the restricted file of the 2012 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 4,694 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) that represents one eligible community that completed the survey. Also included in the data file are 295 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 and social work 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 September 2012 and February 2013. 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 2012 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 2012 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 104 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 39,779 communities in the sampling frame, 11,690 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 180 (1.8% weighted) residential care communities were invalid or out of business. However, 4,578 communities (44% weighted) 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 4,694 communities, for a weighted response rate (for differential probabilities of selection) of 55.4%. Sample weights

[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.

were adjusted to add up to the estimated number of eligible residential care communities (22,200).

Differences in the number of residential care communities between 2010 and 2012

The estimate of the number of residential care community providers varied between the 2010 NSRCF and the 2012 NSLTCP. NCHS continues to take an in-depth look at these differences. Preliminary assessments indicate that the differences in estimates largely stemmed from the eligibility differences between the 2010 NSRCF and the 2012 NSLTCP. 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 low eligibility rate among small sized residential care communities had a large effect on the differences in the eligibility rate for the two surveys.

There could be several reasons for these differences in eligibility rate between the two surveys. One possible reason is that residential care community regulations vary by state and facility bed size, and a larger sample in the 2012 NSLTCP may have more accurately captured whether residential care communities met the eligibility requirements of the study. This may be particularly true in census states where all providers in the state were sampled, since the vast majority of residential care communities are small. However, a more plausible reason that led to eligibility differences may be found in 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), as well as 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. Limited cognitive testing was conducted to assess these eligibility questions, and preliminary findings supported this hypothesis. The other common cause of ineligibility was related to communities exclusively serving severely mentally ill and/or intellectually disabled/developmentally disabled populations. Although during the frame development process, information about residential care communities that exclusively serve these special populations was collected from the state licensing agencies, many state licensing agencies were

[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).

still unable or unwilling to provide listings of these providers. These listings were often maintained at different agencies, and states did not have the manpower to cross-reference the listing. In addition, many state licensing agencies did not provide information on the types of residents served by each provider; therefore, many of these providers could not be eliminated from the states' listings. This issue might have partially accounted for the high percentage of residential care communities that were screened as ineligible on these questions. Since the differences in eligibility were largest in the case of small sized providers, the 2012 estimate of the number of small providers was much lower than the 2010 estimate of the number of small providers. The lower eligibility rate among small sized providers in 2012 also might have explained why the differences in the number of residents between 2010 and 2012 (733,300 versus 713,300) were less notable relative to the difference in the number of providers (31,100 versus 22,200); smaller sized providers account for the majority of communities but house the minority of residents.

In an effort to address these differences, NCHS is working on various measures for the 2014 wave of the NSLTCP, including an in-depth reassessment of regulations (compared to our eligibility definitions) in selected states and revision of eligibility questions based on cognitive testing. NCHS will continue to explore other measures to better understand these differences and address them accordingly.

Questionnaire

The PDF for the residential care community questionnaire of the 2012 NSLTCP is available at: http://www.cdc.gov/nchs/nsltcp/nsltcp_questionnaires.htm. 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 instance, the questionnaire does not have a skip instruction between Question 14 (MEDICAID) and Question 15 (MEDPAID). The intent of Question 15 is to measure the number of residents receiving long-term care services that are financed by Medicaid through direct payment to the residential care community, as opposed to the number of residents whose care, regardless of where they receive it, is paid by Medicaid. In order to make this intent clear and be able to track changes over time, we decided to implement the skip logic in the 2012 survey data, and will add the skip logic in the future NSLTCP surveys. Thus, in the data file, if communities reported not being certified or otherwise set up to participate in Medicaid (MEDICAID=2), their responses to MEDPAID were coded as "-1=Legitimate skip." Items following a skip pattern in the data file are identified in the data dictionary.

Data dictionary

The data dictionary or codebook is provided as a single file containing five sections in the questionnaire: Background Information; Services Offered; Staff Profile; Resident Profile; and Record Keeping. The "Study Eligibility" section is not included in the data file or the data dictionary. 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 survey were identified by bold, uppercase text and were often followed by an arrow (e.g., in Question 10, the instruction reads "→ **SKIP TO QUESTION 11**"), 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 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 were 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. The hardcopy mail questionnaire also included skip instructions using arrows and upper case bold text.

a. 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 they did not collect/track a given health information or left it unanswered (e.g., Question 42a_1; TRMDINFO="2= No" or "-9= Not Ascertained"), then their responses to whether they have the computerized capability to collect/track this health information (e.g., Question 42a_2; ITMDINFO) were coded as "-1= Legitimate skip."

b. For Question 10 (OTHOWN) and Question 10a (OWNHOSP to OWNOTH), the responses to OTHOWN were edited based on responses to OWNHOSP through OWNOTH. If OTHOWN was left blank or answered as "2=No," but the respondent answered "1=Yes" to OWNHOSP, OWNSNF, OWNHHA, OWNHOS, or OWNADSC, then OTHOWN was recoded to "1= Yes."

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 12). 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 participants 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 30).
    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.

Derived variables

1. Number of full-time equivalents by employee staff type (i.e., Question 26: RNFTE1_RC, LPNFTE1_RC, AIDEFTE1_RC, SOCWFTE1_RC)
    a. These variables were derived and provided in the restricted file. Number of full-time and the number of part-time employees for a given staff type 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. Outliers 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 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; and if the number of FTE social work employees was greater than 99.

2. Hours per resident day, by employee staff type (i.e., RNHPPD, LPNHPPD, AIDEHPPD, SOCWHPPD)
    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 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). For a few cases, AIDEHPPD had values greater than 24); these values were coded as 24.

3. Having a computerized capability to collect/track any health information (ANYIT)
    a. This variable was derived from variables (e.g., ITMEDINFO, ITDEM, ITFUNC, ITORDER, ITVIEW for Question 42), indicating whether the residential care community has the computerized capability to collect/track 19 different types of health information.

4. Having a computerized system that supports electronic health information exchange with physicians or pharmacies (ANYEX)
    a. This variable was derived from ITMD and ITPHARM (Question 43).

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. A set of critical items were identified prior to fieldwork; these items were considered necessary for data analysis as they capture essential information about residential care communities (e.g., information on study eligibility, ownership type, owned by other type of organization, Medicaid participation, number of current residents, number of nursing employees, number of male residents, number of residents diagnosed with Alzheimer's disease and other dementias). When returned hardcopy mail questionnaires had critical items missing, interviewers conducted data retrieval via CATI to retrieve missing data for these items; 1,459 cases were sent to CATI for data retrieval. Of those cases, interviewers obtained missing data for 1,156 cases. For another 90 cases, interviewers obtained answers to some of the missing

data or confirmed that the respondent wanted to leave the items as missing. Combined, this represents about 85% of the data retrieval cases.

Imputed data

In the data file, item nonresponse is coded as "-9= Not ascertained." Missing values for the EATHELP (Question34c), BATHHELP (Question 34e), DXALZ (Question 32a), DXDEP (Question 32d), race-ethnicity (Question 29), sex (Question 30), and age (Question 31) variables 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 4,694 respondents, the percentage of imputed records ranged from 4.5% (210 missing responses) for the categories of sex variable (MALE, FEMALE) to 8.7% (408 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. 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 2012 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 2012 NSLTCP RCC survey. Users who apply SRS techniques to the 2012 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 2012 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 2012 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