# *Nursing Home and Assisted Living Surveys*
# Waves 1-2

## User Guide
Version 3

June 30, 2026

Suggested citation to this user guide:
Maust DT, Wagner J, Marcus SC, Wagner L, and Spetz J. 2025. National Dementia Workforce Study User Guide: Nursing Home and Assisted Living Surveys. Ann Arbor: University of Michigan. Available at www.ndws.org.

We thank Piotr Dworak for his assistance preparing this manual. This guide was prepared with funding from the National Institute on Aging (U54AG084520).

# Table of Contents



**PLEASE READ**

- **All publications using NDWS data should include the following citation:**
  - Maust DT, Wagner J, Zimmerman S, Wagner LM, Marcus SC, White EM, Tocco RS, Rothstein J, Hickey JVTS, Pettit AR, Spetz J. The National Dementia Workforce Study: Overview. *J Am Geriatr Soc* 2025;73:3318–26. https://doi.org/10.1111/jgs.70061.

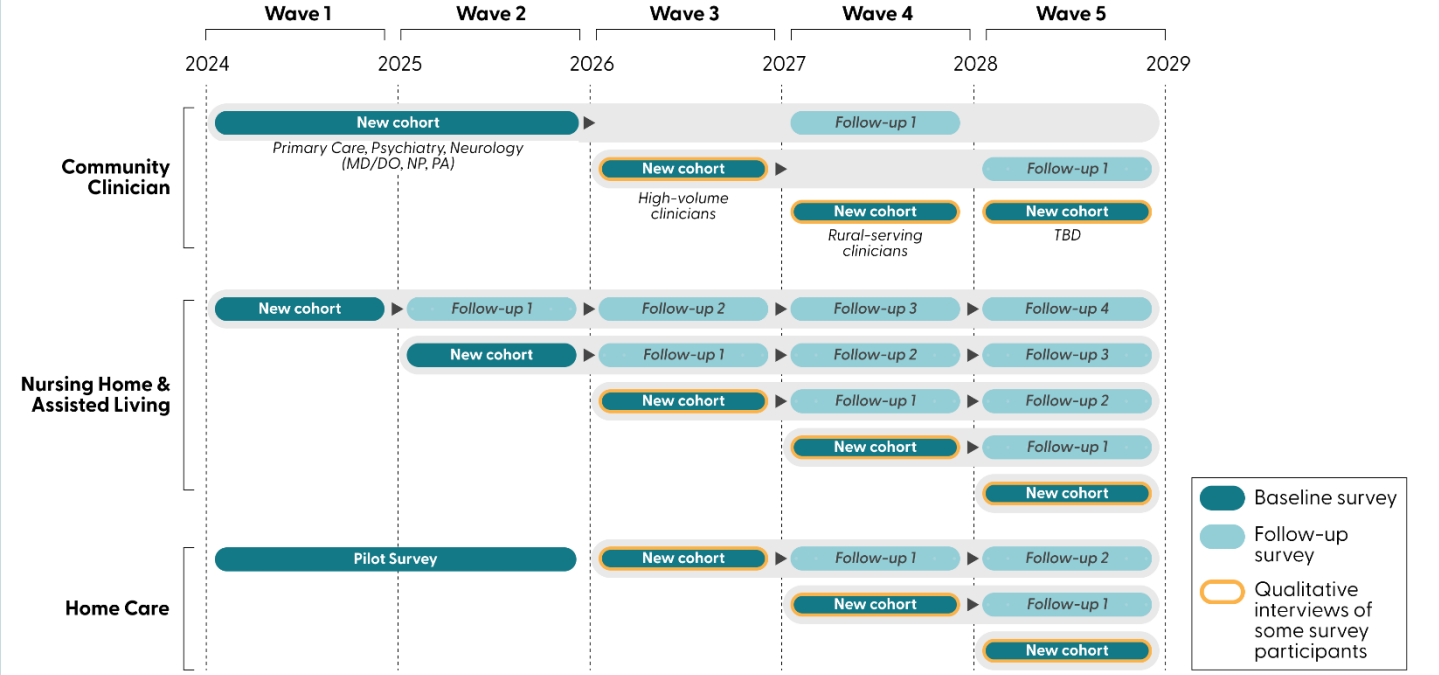
- The appropriate citation for data, if required, is:
  - National Dementia Workforce Study. Produced and distributed by the University of Michigan and the University of California, San Francisco with funding from the National Institute on Aging (Grant U54AG084520), Ann Arbor, MI, www.ndws.org.
    - Publications using the Public Use Files should also include the appropriate doi available on NACDA.

- **All publications should include the following statement in the acknowledgements:** "The National Dementia Workforce Study data is sponsored by the National Institute on Aging (Grant U54AG084520) and is conducted by the University of Michigan and the University of California, San Francisco."

# 1. Introduction and Overview

## 1.1. NDWS Overview

The National Dementia Workforce Study (NDWS), sponsored by the National Institute on Aging (NIA) at the National Institutes of Health (NIH), is a family of nationally representative surveys of the dementia care workforce in the United States. NDWS includes surveys of workers in nursing homes, assisted living communities, home care agencies, and community clinical settings. The following graphic presents an overview of completed (and planned) data collection across the project.




**Please note:** future activities may be modified from those shown here.

This User Guide documents data collection for the **Nursing Home (NH)** and **Assisted Living (AL)** surveys. It provides an overview of the data collection protocol, survey content, information about sampling, and information necessary to analyze the data.

More information about instruments, sampling frames, other data sources available as part of the study, and instructions for accessing all NDWS data, can be found on the study website: NDWS.org.

If you have any questions, please contact info@ndws.org.

## 1.2. Data Access

NDWS survey data are released in two formats:

- **Restricted-use files (RUFs)** available through the NIA-funded LINKAGE platform or the restricted data enclave of the Michigan Center for the Demographic of Aging (MiCDA).
- **Public-use files (PUFs)** available through the National Archive of Computerized Data on Aging (NACDA).

Information about data access is available at: https://www.ndws.org/surveys-and-data/how-to-access-data

Questions may be directed to info@ndws.org.

## 1.3. Who do the NDWS Nursing Home and Assisted Living surveys represent?

### *1.3.1. Nursing Home Survey*

The NH survey represents direct care and nursing staff working in nursing facilities certified by the Centers of Medicare and Medicaid Services (CMS). The facility sampling frame was constructed using CMS administrative sources, including the Minimum Data Set (MDS) and Nursing Home Compare as well as Brown University's LTCFocus and the Payroll-Based Journal (PBJ) records.

- For Wave 1, the final national NH sampling frame included **14,808 nursing homes**, from which **200 facilities** were selected.
- For Wave 2, the updated national NH sampling from included **14,743 nursing homes**, from which **1,000 facilities** were selected.

Within each sampled facility, a roster of eligible staff was obtained to select staff for the survey. Only staff from select direct care roles was eligible. The following roles are used by facilities to report staffing to CMS through the Payroll Based Journal (PBJ). Eligible staff roles included:

- Director of Nursing
- Registered nurse (RN)
- Licensed practical/vocational nurse (LPN/LVN)
- Certified nursing assistant (CNA)
- Nurse aide in training
- Medication aide/technician

### *1.3.2. Assisted Living Survey*

The AL survey represents licensed nurses and direct care workers employed in assisted living communities that provide care to people living with dementia. Unlike NH, AL communities are not federally regulated nor licensed. The AL sampling frame was derived from state licensing information. To select which states to sample in Wave 1, five state replicates—random subsamples of all states—were created from the list of 50 states and District of Columbia (four replicates of 10 states and one replicate with 11 states).

- For Wave 1, we randomly selected two of the replicates, which included 21 states. In Wave 1, across these 21 states, 18,118 AL communities were identified, from which the AL sample of 200 communities was drawn.
- For Wave 2, we updated our state licensing information for all 50 states and the District of Columbia. We identified a total of 39,308 communities, from which a sample of 1,000 was drawn. We also expanded the licensure terms that we used to identify AL communities in each state. This led to the identification of 883 additional AL communities in the 21 states from Wave 1.

Within each sampled community, a roster of eligible staff was obtained to select staff for the survey. While AL communities do not report staffing information to CMS, we used job titles that would be consistent with staff eligible for the NH Staff survey, including:

- Head nurse or Director of Nursing
- Health care supervisor
- Registered nurse (RN)
- Licensed practical/vocational nurse (LPN/LVN)
- Direct care worker

- Certified nurse aide (CNA)/assistant
- Nurse aide/assistant
- Personal care aide/assistant
- Medical technician or medication aide
- Activity staff

# 2. Data Collection Results and Content Documentation

## 2.1. Survey Components

Data collection from all new Wave 2 organizations included two coordinated instruments:

- Administrator Survey, documenting organization-level characteristics
- Staff Survey, documenting workforce respondent characteristics and professional experiences

Administrators also provided rosters of eligible staff, which were used to draw probability samples of staff respondents.

In addition to the new organizations and staff recruited to participate in NDWS, the Wave 2 data collection included the first staff longitudinal follow-up surveys of all staff respondents from Wave 1.

## 2.2. Response Rates by Survey Wave

Table 1. Sample Sizes and Response Rates

| Survey | Sample Size | Completed Surveys | Response Rate |
|---|---|---|---|
| **Wave 1** | | | |
| *Nursing Home* | | | |
| Administrator | 200 | 28 | 14.1%[a] |
| Staff | 722 | 394 | 55.2% |
| *Assisted Living* | | | |
| Administrator | 200 | 35 | 19.4%[a] |
| Staff | 663 | 447 | 67.4% |
| **Wave 2** | | | |
| *Nursing Home* | | | |
| Administrator | 1,000 | 138 | 14.1%[a] |
| Staff: new cohort | 4,240 | 2,374 | 56.0% |
| Staff: Wave 1 follow-up | 394 | 303 | 76.9% |
| *Assisted Living* | | | |
| Administrator | 1,000 | 120 | 12.8%[a] |
| Staff: new cohort | 2,303 | 1,559 | 67.7% |
| Staff: Wave 1 follow-up | 447 | 356 | 79.6% |

[a] In each wave of data collection, some of the units included in the NH and AL organization samples were determined to be ineligible and were excluded from the denominator when calculating the response rate.

## 2.3. Survey Instruments

**Tables 2 and 3** describe the content domains of the data collection instruments used in NDWS, with items described in the order presented to respondents. Where applicable, information on item sources and references is provided.

| Table 2: NDWS Nursing Home and Assisted Living STAFF Survey Content-at-a-Glance[a] | |
|---|---|
| **Section** | **Key Topics Covered** |
| Education, training & experience | Licensure and certification, education level, training experiences, and preparedness for dementia care<br><br>*Includes content from:*<br>• A Minimum Data Set for the Behavioral Health Workforce<br>• NPALS Direct Care Worker Pilot Study, United States Department of Health and Human Services |
| Employment status | Years in long-term care, current employment status, multiple job holding, hours, shifts, supervision, wages, benefits, scheduling, transportation, childcare<br><br>*Includes content from:*<br>• United States Department of Health and Human Services. National Center for Health Statistics. National Nursing Home Survey, 2004. National Nursing Assistant Supplement<br>• Measuring Precarious Work Schedules<br>• NPALS Direct Care Worker Pilot Study, United States Department of Health and Human Services<br>• Pew Research Center, American Trends Panel Wave Survey (Wave 2 Reinterview Addition) |
| Dementia care knowledge, attitudes & practices | Attitudes toward people with dementia, confidence in care tasks, team functioning, environment and safety, PPE, and resource availability and emergency preparedness<br><br>*Includes content from:*<br>• The DAS-6: A Short Form of the Dementia Attitudes Scale<br>• Sense of Competence in Dementia Care Staff (SCIDS) Scale<br>• SOPS ® Workplace Safety Supplemental Item Set for the SOPS Nursing Home Survey |
| Worker outcomes | Job satisfaction, workplace climate, discrimination, harassment, injuries, burnout, retention intentions<br><br>*Includes content from:*<br>• RN4CAST Study<br>• NPALS Direct Care Worker Pilot Study, United States Department of Health and Human Services<br>• NIOSH Worker Well-Being Questionnaire<br>• Development of the Practice Environment Scale of the Nursing Work Index (Wave 2 Addition)<br>• Sheps Center for Health Services Research, University of North Carolina at Chapel Hill<br>• Maslach Burnout Inventory<br>• Relatives of the Impaired Elderly: Correlates of Feelings of Burden (Wave 2 Addition) |
| Demographics | Age, race/ethnicity, immigration, language, disability, household composition, marriage status, caregiving responsibilities, health<br><br>*Includes content from:*<br>• California Board of Registered Nursing 2022 Survey<br>• United States Census Bureau - Citizenship |

| Table 2: NDWS Nursing Home and Assisted Living STAFF Survey Content-at-a-Glance[a] | |
|---|---|
| **Section** | **Key Topics Covered** |
| | • KFF LA Times Survey of Immigrants<br>• Centers for Disease Control and Prevention - Sexual Orientation<br>• United States Census Bureau - Veterans<br>• National Center for Women and Information Technology Demographics Guide<br>• National Health Interview Survey (Wave 2 Addition)<br>• NPALS Direct Care Worker Pilot Study, United States Department of Health and Human Services<br>• NIOSH Worker Well-Being Questionnaire |
| [a] The NH and AL Staff follow-up surveys included questions on employment status, role and compensation changes, organizational and supervisory shifts, job start and end dates, reasons for leaving positions, job search activity, characteristics of new roles, and comparisons between prior and current employment. Full Wave 2 instruments and the crosswalk are available here: https://www.ndws.org/surveys-and-data/surveys | |

| Table 3. NDWS Nursing Home and Assisted Living ADMINISTRATOR Survey Content-at-a-Glance | |
|---|---|
| **Section** | **Key Topics Covered** |
| Organization structure and services | Ownership type, private equity involvement, chain affiliation, years licensed, Medicare/Medicaid certification<br><br>*Includes content from:*<br>• Sheps Center for Health Services Research, University of North Carolina at Chapel Hill |
| Workplace resources and practices | Services, volume of clients/beds, client payer mix, memory care designation, private-pay base rates, EHR adoption<br><br>*Includes content from:*<br>• Sheps Center for Health Services Research, University of North Carolina at Chapel Hill |
| Staffing | Staffing, employee benefits, union representation<br><br>*Includes content from:*<br>• The Office of the National Coordinator for Health Information Technology, Data Brief #39<br>• Sheps Center for Health Services Research, University of North Carolina at Chapel Hill |
| Training practices | New-hire training, continuing education<br><br>*Includes content from:*<br>• Sheps Center for Health Services Research, University of North Carolina at Chapel Hill |

## 2.4. Codebooks

NDWS data are released in two forms: public-use files (PUFs) that have been modified for participant privacy protection (see section 2.5) and restricted-use files (RUFs). The PUFs are available through the National Archive of Computerized Data on Aging (NACDA); RUFs are available through the NIA-funded LINKAGE platform. In the PUF versions, certain data elements are suppressed to mitigate disclosure risk.

For the NH and AL surveys, only the Staff surveys are available as PUFs; the Administrator surveys are only available as RUFs. In addition, **the Wave 2 PUF only includes new staff respondents; the Wave 1 staff follow-up data are only available in the restricted-use enclave**.

Separate data dictionary codebooks are provided for the PUF and RUF versions of each NDWS survey dataset. The PUF codebook is available on NACDA; the RUF version is available on LINKAGE for approved users. For each item, the codebooks include the variable name, label, response options, and a frequency distribution of responses, including special values used to represent missing data. The codebooks are intended to support data interpretation and analytic use of the NDWS survey files.

### 2.5. PUF Preparation and Disclosure Review

Protecting participant privacy is critical from both a compliance and ethics perspective. It must also be balanced with ensuring data utility for research. In accordance with best practices, we offer several layers of privacy protections for the de-identified, NDWS PUFs. First, we removed all direct identifiers under the HIPAA safe harbor method (e.g., date of birth) or variables for which we were concerned that participant identity could be considered readily ascertainable under the Human Subjects Research Regulations (e.g., geographic area). Second, to protect participant privacy while maintaining data utility, we applied an anonymity algorithm that converted continuous variables to categories, collapsing sparse categorical options, and suppressing the minimum number of identifying cells necessary to ensure that every unique combination of indirect identifiers is shared by no fewer than three respondents. Third, we removed all variables alone or in conjunction which we believed posed a potential risk to the participants' financial standing, employability, or reputation as required for exempt research under 45 CFR § § 46.104(d)(2). Additional detail is provided in the PUF documentation for each survey.

### 2.6. Special Values for Missing Data

NDWS survey datasets use standardized special values to represent different forms of missing or inapplicable data in the survey. These special values are documented in the data dictionary codebooks and are reflected in the frequency distributions provided for each survey item. **Table 4** below describes the special values used across NDWS surveys, and the example illustrates how these values appear in a codebook frequency table for an individual survey item.

Table 4. Special Values for Missing Data

| Missing value | Represents |
|---|---|
| "." Or " " | The item was not displayed for this type of respondent (e.g., only Nurse Practitioners saw item *Field*; other respondents would have ".") |
| -9 | The item was displayed but they did not provide answer |
| -8 | Respondents selected "don't know" |
| -7 | Respondents provided an out-of-range value (the web version did not allow this; only possible where the respondent used a paper instrument) |
| -5 | Legitimate skip |

The following example shows the distribution of the *LicenseNowRN* variable available in the NH Staff survey (Restricted Use File).

**LicenseNowRN**

**Label:** Currently holds RN license

**Type:** numeric | **Length:** 8

| Frequencies for LicenseNowRN | | |
|---|---|---|
| **Value** | **Count** | **Percent** |
| -9=Refusal | 2 | 0.55 |
| 0=not selected | 309 | 85.59 |
| 1=selected | 50 | 13.85 |
| Missing | 33 | . |

# 3. Sampling

## 3.1. Overview

Both NH and AL surveys employed multistage, stratified probability sampling designs. Final survey weights incorporate selection probabilities, adjustments for nonresponse, and poststratification factors. **Survey weights, strata, and cluster identifiers must be used in all analyses.**

## 3.2. NH Facility Sampling

NH facilities were sampled using stratified random sampling based on:

- Number of certified beds (tertiles)
- Rural versus urban location (derived from RUCA codes)
- Percentage of Medicaid residents

Bed tertiles were defined as 1-80, 61-120, and 121+ certified beds. The rural/urban status was assigned by first geocoding the addresses then the Rural-Urban Commuting Area (RUCA) code of the Census Tract of the NH was added. RUCA codes 7–10 designated as “rural.” Ratio of Medicaid residents was set at a median percentage (64.4%) of residents that pay via Medicaid.

In Wave 1, a sample of 200 facilities was allocated to the strata based upon the proportion of staff estimated to be in the stratum. There was an additional requirement that there be at least six nursing homes allocated to each stratum (**Table 5**).

Table 5. Distribution of Facilities, Estimated Staff, and Allocated Sample by Stratum (Wave 1)

| Stratum | Bed tertile[a] | Medicaid Low/High | Rural[c] | Facility Count | Est. Total Staff, N | Est. Staff in Stratum, % | Facilities in Stratum, % | Sampled NH |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | below | no | 2338 | 47,579 | 9.4% | 15.8% | 20 |
| 2 | 1 | below | yes | 861 | 13,876 | 2.7% | 5.8% | 6 |
| 3 | 1 | above | no | 1317 | 25,921 | 5.1% | 8.9% | 10 |
| 4 | 1 | above | yes | 612 | 10,349 | 2.0% | 4.1% | 6 |

| Table 5. Distribution of Facilities, Estimated Staff, and Allocated Sample by Stratum (Wave 1) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | 2 | below | no | 2261 | 80,788 | 16.0% | 15.3% | 32 |
| 6 | 2 | above | no | 2295 | 74,548 | 14.7% | 15.5% | 30 |
| 7 | 3 | below | no | 1701 | 95,338 | 18.8% | 11.5% | 38 |
| 8 | 3 | above | no | 2350 | 125,414 | 24.8% | 15.9% | 44 |
| 9 | 2-3 | both | yes | 1073 | 32,120 | 6.3% | 7.2% | 14 |
| | | | | **14,808** | **505,933** | **100%** | **100%** | **200** |

[a] certified bed tertiles: 1 (1-80 beds), 2 (61-120), and 3 (121+)
[b] below or above the median percentage (64.4%) of facility residents that pay via Medicaid
[c] defined as facility addresses with a Rural–Urban Commuting Area (RUCA) code of 7-10

In Wave 2, a sample of 1,000 facilities was selected by allocating to the proportion of estimated staff each stratum with the additional requirement that there be at least six nursing homes allocated to each stratum (**Table 6**).

| Table 6. Distribution of Facilities, Estimated Staff, and Allocated Sample by Stratum (Wave 2) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Stratum** | **Bed tertile[a]** | **Medicaid Low/High[b]** | **Rural[c]** | **Facility Count** | **Est. Total Staff, N** | **Est. Staff in Stratum, %** | **Facilities in Stratum, %** | **Sampled NH** |
| 1 | 1 | below | no | 2,330 | 48,455 | 9.4% | 15.8% | 134 |
| 2 | 1 | below | yes | 860 | 13,819 | 2.7% | 5.8% | 96 |
| 3 | 1 | above | no | 1,302 | 26,059 | 5.1% | 8.8% | 76 |
| 4 | 1 | above | yes | 601 | 10,195 | 2.0% | 4.1% | 68 |
| 5 | 2 | below | no | 2,266 | 82,338 | 16.1% | 15.4% | 132 |
| 6 | 2 | above | no | 2,282 | 75,707 | 14.8% | 15.5% | 136 |
| 7 | 3 | below | no | 1,698 | 97,052 | 18.9% | 11.5% | 98 |
| 8 | 3 | above | no | 2,337 | 126,917 | 24.7% | 15.9% | 136 |
| 9 | 2-3 | both | yes | 1,067 | 32,345 | 6.3% | 7.2% | 124 |
| | | | | **14,743** | **512,888** | **100%** | **100%** | **1,000** |

[a] certified bed tertiles: 1 (1-80), 2 (61-120), and 3 (121+)
[b] below or above the median percentage (64.4%) of facility residents that pay via Medicaid
[c] defined as facility addresses with a Rural–Urban Commuting Area (RUCA) code of 7-10

Additional details about sampling and weighting for each wave of the NH survey will be released as a supplement.

## 3.3. AL Community Sampling and Weighting

AL sampling for Wave 1 occurred in three stages:

1) Selection of states
2) Selection of AL communities within states
3) Selection of staff within sampled AL communities

**In Step 1,** we created a list of states with several attributes. We sorted that list by Total Population 65+ from the 2021 American Community Survey Public Use Microdata Sample file. There were no ties once the states were sorted by that variable. We then created 5 random replicates of states by randomly choosing an integer between 1 and 5 and then numbering the states starting from that random start. After the 5$^{th}$ number was counted, we restarted at 1. This created 5 replicates (4 with 10 states and 1 with 11 states since DC was included. We took two of these replicates as a sample of states. The resulting first-stage selection probability, therefore, can be calculated as:

$\pi_0 = \frac{c}{C} = \frac{2}{5}$,

where $c$ is the number of selected replicates and $C$ is the total population of replicates. This selection probability is denoted with a 0 to avoid confusion with the first-stage notation used for the nursing home and home care agency surveys (i.e. the organization).

**In step 2,** we contacted the 21 states selected for the Wave 1 sample and obtained lists of AL communities from each of those states. The resulting list had a total of **18,211** AL communities (**Table 7**). Each state made a different set of administrative variables available. Therefore, we specified a minimum set of variables as contact information (street address, telephone number, and an email address) as well as the number of certified beds, which we used as a proxy measure of community size. (AL communities with fewer than 4 certified beds [n=93] were removed from the sample frame, following the approach used in the National Post-acute and Long-term Care Study). We used street address to geocode communities, which allowed us to assign to a Census Block and then classify rurality (as with the NH survey, RUCA codes we defined as "rural").

AL communities were stratified by:

- the number of certified beds,
- rural location, and
- Census Region

Number of certified beds was initially divided into tertiles of 4-6 certified beds, 7-34 bed, and 35+ certified beds. We then added additional category by splitting the 35+ group into 35-99 and 100+ to account for the fact that in some states there is a regulatory limit set at 100 beds and that more than 80% of all beds were in the 35+ tertile. AL communities with RUCA codes 7-10 were identified as rural; only ~4% of all AL communities were rural. Therefore, defined a separate stratum for rural AL communities, merging all rural AL communities regardless of size or Census Region into a single stratum. Finally, there were few AL communities in the Northeast, so we combined the Northeast and Midwest Regions.

Table 7. Distribution of Communities and Beds by Stratum

| Stratum | Bed class | Census region | Communities, N | Communities, % | Certified Beds, N | Beds, % | Sampled AL |
|---|---|---|---|---|---|---|---|
| 1 | 1 to 6 | Midwest-Northeast | 1,215 | 6.7% | 5,767 | 0.9% | 4 |
| 2 | | South | 1,456 | 8.0% | 8,554 | 1.3% | 4 |
| 3 | | West | 6,417 | 35.4% | 37,862 | 5.7% | 11 |
| 4 | 7 to 34 | Midwest-Northeast | 669 | 3.7% | 13,215 | 2.0% | 4 |
| 5 | | South | 1,058 | 5.8% | 16,581 | 2.5% | 5 |
| 6 | | West | 864 | 4.8% | 12,399 | 1.9% | 4 |
| 7 | 35 to 99 | Midwest-Northeast | 1,383 | 7.6% | 89,354 | 13.4% | 26 |
| 8 | | South | 1,372 | 7.6% | 89,861 | 13.4% | 26 |
| 9 | | West | 643 | 3.5% | 43,147 | 6.5% | 12 |
| 10 | 100+ | Midwest-Northeast | 801 | 4.4% | 111,129 | 16.6% | 32 |
| 11 | | South | 788 | 4.3% | 106,547 | 15.9% | 32 |
| 12 | | West | 690 | 3.8% | 107,652 | 16.1% | 32 |
| 13 | n/a | Rural | 762 | 4.2% | 26,739 | 4.0% | 8 |

| Table 7. Distribution of Communities and Beds by Stratum | | | | | |
|---|---|---|---|---|---|
| | **18,118** | **100%** | **668,808** | **100%** | **200** |

For **Wave 2**, we expanded the frame to include all 50 states and the District of Columbia. Further, we expanded the licensure terms used to define "assisted living." These changes led to a frame of 39,308 AL communities (see **Table 8**). As a result $\pi_0$ is equal to 1 for all states and DC.

Also for Wave 2, there was an interest in adding oversampling of underserved populations in addition to rural areas. This was formalized using the Area Deprivation Index (ADI; Kind AJH, Buckingham W. DOI: 10.1056/NEJMp1802313). A common approach to using the ADI is to focus on the top decile of this measure (91-100$^{th}$ percentiles, i.e. the most disadvantaged areas). Overall, about 4% of AL communities are in areas of high ADI. Given this relatively small size, we decided to treat this as a separate stratum (stratum 14). There was some overlap with AL in rural areas and ADI. About 25% of the High ADI cases were also in rural areas (about 1% of the total frame). We included these in the High ADI stratum.

| Table 8. Distribution of Communities and Beds by Stratum | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Stratum** | **Bed class** | **Census region** | **Communities, N** | **Communities, %** | **Certified Beds, N** | **Beds, %** | **Sampled AL** |
| 1 | 1 to 6 | Midwest-Northeast | 2883 | 7.3% | 12,770.0 | 0.9% | 18 |
| 2 | | South | 3123 | 7.9% | 17,343.0 | 1.2% | 19 |
| 3 | | West | 7705 | 19.6% | 44,235.0 | 3.1% | 54 |
| 4 | 7 to 34 | Midwest-Northeast | 2705 | 6.9% | 49,300.9 | 3.4% | 18 |
| 5 | | South | 3077 | 7.8% | 45,686.4 | 3.2% | 18 |
| 6 | | West | 3146 | 8.0% | 40,609.0 | 2.8% | 18 |
| 7 | 35 to 99 | Midwest-Northeast | 3642 | 9.3% | 229,129.5 | 15.9% | 111 |
| 8 | | South | 2768 | 7.0% | 179,830.3 | 12.5% | 101 |
| 9 | | West | 1745 | 4.4% | 113,591.0 | 7.9% | 56 |
| 10 | 100+ | Midwest-Northeast | 1765 | 4.5% | 248,002.8 | 17.2% | 144 |
| 11 | | South | 1314 | 3.3% | 172,874.5 | 12.0% | 134 |
| 12 | | West | 1106 | 2.8% | 161,549.5 | 11.2% | 144 |
| 13 | n/a | Rural and Low ADI | 2730 | 6.9% | 79,038.6 | 5.5% | 77 |
| 14 | n/a | High ADI | 1599 | 4.1% | 49,688.1 | 3.4% | 88 |
| | | | 39,308 | 100.0% | 1,443,649 | 100.0% | 1,000 |

Additional details about sampling and weighting for each wave of the AL survey will be released as a supplement.

### 3.4. Weights to use when combing Data from Wave 1 and Wave 2

It is possible to combine data from Waves 1 and 2. If the purpose is to compare estimates from Waves 1 and 2, then this should be accomplished using the existing wave-specific weights as this is a comparison of two different populations (i.e. the population of AL or NH staff at Wave 1 and the population of AL or NH staff at Wave 2).

However, combining the data in order to increase the sample size requires an assumption of a stable estimand. That is, it is necessary to assume that the quantity being estimated is stable over the two-year period. In order to combine the samples, changes to the weights were made and a new variable (`FinalWeight_W01_W02`) should be used.

To create the appropriate weight for combining Wave 1 and Wave 2, first, the probabilities of selection were modified to account for the increased sampling rate. This impacted the NH and AL facility probability of selection. It did not impact the staff selection as samples of staff were only drawn at one point in time. Second, we calculated new poststratification adjustments for NH and AL facilities. We did not recalculate nonresponse adjustments. These adjustments scale the NH and AL facilities and staff to the size of the sample as described in detailed sampling supplement referenced below. These adjustments are specific to the Wave of data collection. As such, we did not re-estimate a single overall nonresponse adjustment. This is analogous—but not equivalent—to having a single model with an indicator variable for Wave 2 that is also interacted with every predictor.

The variable `FinalWeight_W01_W02` to use in analyses combining baseline data from Waves 1 and 2 is available as supplemental RUF files for the NH and AL Staff surveys (ndws_w1w2_nh_wgts and ndws_w1w2_al_wgts in SAS and Stata versions) along with the survey data. The supplemental files include the following:

Table 9. Variables in the supplemental file to combine W1 and W2 Staff baseline surveys

| **Variable:** | **Variable Label:** |
|---|---|
| `NonresponseAdjustment_W01_W02` | First-Stage Nonresponse Adjustment (Wave1+Wave2) |
| `StaffNonresponseAdjustment_W01_W02` | Second-Stage Nonresponse Adjustment (Wave1+Wave2) |
| `StaffSamplingWeight_W01_W02` | Second-Stage Selection Weight (Wave1+Wave2) |
| `FacilitySelectionProb_W01_W02` | First-Stage Selection Probability (Wave1+Wave2) |
| `FacilitySelectionWeight_W01_W02` | First-Stage Selection Weight (Wave1+Wave2) |
| `FacilityFinalWeight_W01_W02` | Final facility weight (Wave1+Wave2) |
| `Stratum_W01_W02` | Sampling Stratum |
| `FinalWeight_W01_W02` | Final Weight Combining Selection Weights, Nonresponse Adjustments, and Poststratification Factors (Wave 1 + Wave 2) |

# 4. Example Code for Weighted Analysis and Variance Estimation

The Nursing Home and Assisted Living Staff surveys are stratified cluster samples of staff. There were two stages of sampling. In the first stage, nursing homes and assisted living communities were selected from a stratified sampling frame. These organizations are the clusters. In a second stage of sampling, a sample of staff members were drawn. After data collection, a series of weighting steps were implemented (described in the Sampling and Weighting Supplement). These include sample selection weights, nonresponse adjustments, and poststratification. The result was a final weight variable that should be used in all analyses.

Each data release includes two weights for analysis of the new baseline Staff surveys: one weight to use when only data from a given new staff Wave is used, and a second weight to use when all new respondents from the current and prior waves are combined (e.g., in the Wave 2 release, this will be the weight to use if combining all new staff surveys from Waves 1 and 2).

Finally, the Wave 2 data release includes follow-up of the Wave 1 staff respondents. That file includes a separate weight to use in the analysis of those follow-up surveys, which accounts for patterns of non-response among those staff who participated in Wave 1 and were invited to participate in Wave 2.

**To obtain correct standard errors and confidence intervals in statistical analyses, the sample design, including clustering, stratification, and survey weights, must be specified in statistical analysis software. Failing to account for the design can lead to incorrect inferences.**

## 4.1. SAS

This example demonstrates estimation of a mean and its design-adjusted standard error using PROC SURVEYMEANS, with stratification specified via the STRATA statement, the cluster specified with the CLUSTER statement, and survey weights specified via the WEIGHT statement. For baseline analyses of a single wave of the NH or AL Staff surveys, the stratification variable is `stratum` , the cluster variable is `NDWS_FACID` (RUF version) or `pSampleCluster` (PUF version), and the survey weight is `FinalWeight`. An analysis that combines multiple baseline waves of staff surveys uses the same stratification and cluster variables but should use the survey weight variable `FinalWeight_W01_W02` (described above in section 3.4).

As noted earlier, NDWS survey data are released in two forms: a restricted use file (RUF) available through the LINKAGE platform and a public use file (PUF) with certain elements removed to protect confidentiality. The RUF includes all original design variables, while in the PUF the variables may have been altered for disclosure protection.

The following SAS code will generate a weighted estimate for the level of satisfaction with aspects of the current job, in this case the overall job satisfaction (variable name: `JobOverall`) rated on the scale from 1 to 5 where 1 = very dissatisfied and 5 = very satisfied. The estimate will incorporate all sample design elements

**Table 9** provides sample code using the Restricted Use File (RUF) data set and its Public Use File (PUF) equivalent. The only difference between the PUF and RUF examples are that the RUF data uses facility ID (`NDWS_FACID`) as clusters and the PUF dataset uses a `pSampleCluster` variable for the same purpose; the analyses yield the same results.

| Table 9. Example Code for Weighted Analysis | |
|---|---|
| **Restricted Use File (RUF)** | **Public Use File (PUF)** |

```
PROC SURVEYMEANS DATA=AL_Staff_RUF;
  STRATA stratum;
  CLUSTER NDWS_FACID;
  WEIGHT FinalWeight;
  VAR JobOverall;
RUN;
```

| Data Summary | |
|---|---|
| Number of Strata | 8 |
| Number of Clusters | 35 |
| Number of Observations | 447 |
| Sum of Weights | 738856.815 |

| Statistics | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Label | N | Mean | Std Error of Mean | 95% CL for Mean | |
| JobOverall | Level of satisfaction: Overall job | 447 | 3.409429 | 0.075981 | 3.25352950 | 3.56532800 |

```
PROC SURVEYMEANS DATA=AL_Staff_PUF;
  STRATA stratum;
  CLUSTER pSampleCluster;
  WEIGHT FinalWeight;
  VAR JobOverall;
RUN;
```

| Data Summary | |
|---|---|
| Number of Strata | 8 |
| Number of Clusters | 35 |
| Number of Observations | 447 |
| Sum of Weights | 738856.815 |

| Statistics | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Label | N | Mean | Std Error of Mean | 95% CL for Mean | |
| JobOverall | Level of satisfaction: Overall job | 447 | 3.409429 | 0.075981 | 3.25352950 | 3.56532800 |

Similar survey procedures are available in SAS for other common analyses, including cross-tabulations (PROC SURVEYFREQ), linear regression (PROC SURVEYREG), and logistic regression (PROC SURVEYLOGISTIC).

## 4.2. Stata

In Stata, analyses must first declare the survey design. For the Nursing Home or Assisted Living Staff surveys using RUF data, this includes the weight, stratification, and cluster variables as follows:

```
svyset [pweight=FinalWeight], strata(stratum), psu(NDWS_FACID).
```

- Weight is declared via `pweight=FinalWeight`
- Stratification via `strata(stratum)`
- Cluster via:
  - RUF: `psu(NDWS_FACID)`
  - PUF: `psu(pSampleCluster)`

Then, it is necessary to reference the survey design using the `svy` prefix command. For example:

```
svy: mean JobOverall
```

# 5. History of Changes

Wave 2 data release:

- Version 2 (May 26, 2026):
  1. Corrects Table 9 to appropriately label the RUF and PUF example code

- Version 3 (June 30, 2026):
  1. Adds citation/acknowledgment information
  2. Adds study overview graphic
  3. Updates number of follow-up respondents in the Nursing Home Staff survey (Table 1)
  4. Adds new missing data value to Table 4
  5. Includes information about weights to combine baseline waves of data (section 3.4)
  6. Updates to section 4.2