Community Clinician Survey
Wave 1

User Guide

February 2026

Suggested Citation:
Maust DT, Wagner J, Marcus SC, Wagner L, and Spetz J. 2025. National Dementia Workforce Study User Guide: Community Clinician Survey. 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

1. Introduction and Overview

1.1. NDWS Overview

The National Dementia Workforce Study, sponsored by the National Institute on Aging (NIA) of the National Institutes of Health (NIH), is comprised of a family of surveys of the dementia care workforce in the United States. This User Guide addresses the Community Clinician survey; NDWS also includes Nursing Home, Assisted Living, and Home Care surveys, which are addressed elsewhere.

This document describes Wave 1 NDWS data collection, which spanned the period between August 2024 and April 2025.

This User Guide accompanies the release of the NDWS survey data. Restricted-access data are available on the NIA-funded LINKAGE platform, while the public use files (PUFs) are available through the National Archive of Computerized Data on Aging (NACDA). Information about accessing data is available here: https://www.ndws.org/surveys-and-data/how-to-access-data

The User Guide provides an overview of the data collection protocol, information about sampling and weighting, and all 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. Who does the NDWS Community Clinician survey represent?

The Community Clinician survey is drawn from a random sample of clinicians from all 50 states and DC. It includes primary care physicians (including geriatricians), primary care nurse practitioners, non-surgical physician assistants, psychiatrists, psychiatric-mental health nurse practitioners, and neurologists who provided outpatient care (including residential settings) or prescribed medication to Medicare beneficiaries with a dementia diagnosis in the U.S. in federal fiscal year 2023 (i.e., October 2022 through September 2023).

2. Content Documentation

2.1. Community Clinician Survey

The Community Clinician (CC) sample was derived from a comprehensive sampling frame of **492,186** eligible providers, identified via national Medicare outpatient and professional claims data. From this national cohort of clinicians providing care to Medicare beneficiaries, a representative sample of **25,000** was systematically selected for the study.

Between August 2024 and April 2025, the sample of 25,000 clinicians was invited to complete a 25-minute survey through a series of mailed and emailed invitations and reminders. All clinicians received both a web-based survey invitation and a paper-and-pencil (PAPI) version of the NDWS survey.

2.2. Wave 1 Data Collection

Data collection outcomes are summarized below.

Table 1. Survey Completion and Sample Sizes			
Survey	**Sample Size**	**Number of Completed Surveys**	**Survey Completion Rate**
Community Clinician	25,000	4,699	18.9%

2.3. Data Collection Instruments

Table 2 describes 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 Community Clinician Survey Content-at-a-Glance	
Section	**Key Topics Covered**
Education, Training & Experience	Licensure, education, specialty training, board certification, preparedness for dementia care, years in practice Includes content from: • Final MDS: Physicians. Federation for State Medical Boards
Employment Status	Number of clinical and non-clinical jobs, non-clinical roles (e.g., research, teaching)
Practice Settings & Characteristics	Practice setting, supervision, hours, staffing mix, team composition, EHR use, geographic location
Patient Panel & Scheduling	Panel size, dementia prevalence and severity, visit volume and length, caregiver involvement, interpreter services
Processes of Care: Dementia Screening, Diagnosis and Management	Screening tools, diagnostic confidence, referrals, biomarkers, medications, care priorities, community resources, barriers
Job Outcomes	Job satisfaction, burnout, intent to leave position Includes content from: • Maslach Burnout Inventory
Demographics	Age, race/ethnicity, language, household composition, caregiving responsibilities, health Includes content from: • KFF LA Times Survey of Immigrants • California Board of Registered Nursing 2022 Survey • NIOSH Worker Well-Being Questionnaire

2.4. Codebooks

NDWS data are released in two forms: public-use files (PUFs) that have been modified for participant privacy protection (see section 2.3) 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.

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.

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.

2.5. 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 3** 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 3. 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)

The following example shows the distribution of the *FellowDidNotFinish* variable available in the Community Clinician Survey (Restricted Use File).

FellowDidNotFinish

Label: Fellowship training: Did not complete a fellowship

Type: numeric | **Length:** 8

Value	Count	Percent
-9=Refuse	144	7.17
0=not selected	571	28.43
1=selected	1293	64.39
Missing	2691	.

3. Sampling and Weighting

3.1. Overview

This section describes the sampling and weighting for the NDWS Community Clinician (CC) survey. It describes the procedures for constructing the sampling frame, implementing stratification, and selecting the sample, followed by the procedures for applying nonresponse adjustments and poststratification factors used in developing the final weighting adjustments. The resulting weights, along with stratum and cluster variables, must be included in all statistical analyses to correctly estimate sampling variance, and ensure valid statistical testing.

3.2. Sampling Frame

The sampling frame was constructed using national Medicare outpatient and professional claims data to identify clinicians who provided care to Medicare beneficiaries with a recorded dementia diagnosis. From these claims, National Provider Identifiers (NPIs) were extracted for eligible clinicians based on their recorded licensure and specialty. NPIs were then linked to the National Plan and Provider Enumeration System (NPPES) to obtain provider contact information and demographics. This process yielded a total of 492,186 eligible clinicians.

In addition to the clinician type (licensure and specialty), we also stored several variables determined from the CMS data, including:

- Number of patients with dementia cared for by the clinician
- Number of low-income patients (i.e., dually-eligible for Medicaid and/or the Part D low-income subsidy [LIS])
- The setting of a clinician's patient care encounters (i.e., did they practice in outpatient and/or residential settings, or were they identified only through prescription claims in Medicare Part D)
- Race/ethnicity composition of patients with dementia for each clinician's panel

Since the clinician practice addresses in NPPES may be outdated, we updated address information using a commercial vendor. We then geocoded the updated addresses and added additional variables about the characteristics of the area surrounding the practice. An example of an added variable was urbanicity, classified according to Rural–Urban Commuting Area (RUCA) codes, with RUCA codes 7–10 designated as "rural." More detailed information about CC sample frame construction is available in the "NDWS Wave 1 Sample Frame" documentation available on NDWS.org

We used five stratification variables. Three variables were used to create explicit strata to inform sampling:

- Clinician type (license and specialty)
- Urban/rural status of the clinician
- Whether the clinician has more/less than the overall median number of low-income patients with dementia in their panel

Certain clinician types (categories 4–6) had such a low number of rural cases that we collapsed these with the urban counterparts. **Table 4** presents the strata and the corresponding counts of clinicians from the sampling frame within each stratum.

Table 4. Community Clinician Stratification

Stratum	Clinician Type	Low-income	Rural	Frame Count	Frame Percent
1	Primary care physician	No	No	113,442	23.05
2			Yes	6,712	1.36
3		Yes	No	72,048	14.64

Table 4. Community Clinician Stratification					
4			Yes	5,234	1.06
5	Primary care NP	No	No	109,516	22.25
6			Yes	9,863	2.00
7		Yes	No	41,717	8.48
8			Yes	2,651	0.54
9	PA	No	No	65,613	13.33
10			Yes	3,841	0.78
11		Yes	No	14,062	2.86
12			Yes	759	0.15
13	Psychiatrist	No	Both	14,764	3.00
14		Yes	Both	4,799	0.98
15	Psychiatric-mental health NP	No	Both	8,400	1.71
16		Yes	Both	3,423	0.70
17	Neurologist	No	Both	5,391	1.10
18		Yes	Both	9,951	2.02
NP: nurse practitioner; PA: physician assistant					

Two other variables were used to create **implicit strata**:

- Setting of care (any residential, outpatient only, Part D only)
- Number of patients with dementia cared for by the clinician

"Implicit strata" were used as part of the systematic sampling procedure described in section 4.4.

3.3. Allocation

We set target numbers of respondents for each of the clinician types. Given the relatively small population sizes of psychiatric-mental health nurse practitioners (Psych NPs), psychiatrists, and neurologists, we oversampled clinicians from those categories. **Table 5** summarizes the distribution of clinicians in the sampling frame and the planned sample allocation by clinician type. Columns report the percentage of clinicians in the frame, the planned target sample size, the sample percent and the corresponding relative sampling rate (sample percent divided by frame percent). Sample weights will be constructed for use in analyses to account for differential sampling probabilities across clinician types.

Table 5. Sample and Sample Allocation by Clinician Type				
Clinician Type	**Frame Percent**	**Planned Sample Size**	**Sampled Percent**	**Relative Sampling Rate**
Primary care physician	40.11	7,925	31.70	0.79
Primary care NP	33.27	6,575	26.30	0.79
PA	17.12	3,750	15.00	0.88
Psychiatrist	3.97	2,250	9.00	2.27
Psychiatric-mental health NP	2.40	2,250	9.00	3.75
Neurologist	3.12	2,250	9.00	2.88
Total	**100%**	**25,000**	**100%**	

3.4. Sample Selection

The sample was selected using systematic selection to add implicit stratification. The list of clinicians was sorted within strata by:

- Rural status (collapsed for psychiatrists, psychiatric-mental health NPs, and neurologists)
- Setting of care
- Binary indicator for above/below median number of dementia patients

Systematic selections (every k^{th} selection starting from a random start between 1 and k) were made by stratum using the targeted sample sizes for each stratum as shown above in **Table 5**. This results in the following sample selection equation:

$$\pi_h = \frac{n_h}{N_h}$$

Where:
h = stratum index
n_h = sample size allocated to each stratum
N_h = size of each stratum

All units within each stratum share the same probability of selection. The following is the formula for the corresponding sample selection weight:

$$w_h = \frac{1}{\pi_h}$$

3.5. Nonresponse Adjustment

We evaluated nonresponse patterns and implemented weighting adjustments to ensure the integrity of the survey findings. Of the 25,000 sampled clinicians, 102 were identified as ineligible due to retirement or departure from the medical field. The overall response rate for the Community Clinician survey was calculated as 4,699 / (25,000 − 102) = 18.9%. This response rate was calculated in accordance with standards established by the American Association for Public Opinion Research (AAPOR). Specifically, this corresponds to the AAPOR RR2 definition which assumes that all cases of unknown eligibility are in fact eligible and therefore represents the most conservative approach to estimating the response rate. **Table 6** presents the response rates for selected key subgroups.

Table 6. Response Rates by Sample Frame Subgroups

Variable	Value	Response rate range
Clinician type	Primary care physician	17.14%–21.23%
	Primary care NP	
	PA	
	Psychiatrist	
	Psychiatric-mental health NP	
	Neurologist	
Number of low-income patients with dementia	≤10	16.64%–19.69%
	11-20	
	>20	
Number of patients with dementia	≤10	16.79%–19.75%
	11-20	
	21-50	
	>50	
Census region	Midwest	17.51%–21.26%
	Northeast	
	South	
	West	
Setting	Any residential	17.27%–19.57%
	Outpatient (non-residential)	
	Part D prescriber only	
Provider sex (from NPPES)	Female	17.80%–19.54%
	Male	

Table 6. Response Rates by Sample Frame Subgroups		
Number of patients with dementia who are Black or Hispanic	≤10	14.4%-19.72%
	11-20	
	>20	
Sole proprietor (from NPPES)	No	17.65%–19.19%
	Yes	

Consistent with guidance from Little and Vartivarian (2005), who suggest that variables used in nonresponse adjustment should be correlated with survey variables, a two-step variable selection and modeling strategy was used to support development of weights for nonresponse adjustment. In the first step, LASSO regression was applied among survey respondents to identify frame variables (**Table 6**) associated with each of 12 selected key survey items (**Table 7**), to reduce bias and improve the accuracy of survey estimates. In the second step, the frame variables identified in this process were used as predictors in response-propensity models estimated for the full sample (respondents and nonrespondents), forming the basis for nonresponse adjustment. The key survey variables are listed in the table below.

Table 7. Key Survey Variables Used in Nonresponse Models

Question Text	**Variable Name(s)**	**Variable Type**	**Answer Options**
To what extent has your formal training prepared you to provide care to people with dementia?	TrainPrepare	Binary	• Adequately prepared • Not adequately prepared
How many Full-time years have you been practicing as a physician, physician assistant or nurse Practitioner (your current license)?	PracticeFT	Continuous	Full-time years practicing
How many paid clinical jobs do you have?	JobsClinical	Binary	• Two or more • Not
Do you have any other non-clinical paid jobs?	JobsNonClinical	Binary	• Yes • No
In a typical week, how many hours do you usually work in your principal clinical job?	JobHoursWeek	Continuous	Hours worked per week
Are you in a supervisory or management role in your principal clinical job?	JobSupervise	Binary	• Yes • No
How many years have you been working for your current employer?	JobYears	Continuous	Years with current employer
In your practice, which best describes the extent to which patient care activities are documented in an EHR?	EHR	Binary	• Fully electronic • Not fully electronic
How satisfied are you with aspects of your principal clinical job?	SatisfiedTime SatisfiedLoad SatisfiedSchedule SatisfiedAutonomy SatisfiedSalary SatisfiedDev SatisfiedRespect SatisfiedAdmin SatisfiedInput	Continuous	Average score across all items
To what extent do you feel confident diagnosing dementia and mild cognitive impairment?	DiagDemUnder65 DiagDem65 DiagMildUnder65 DiagMild65	Continuous	Average score across all items
Thinking of the care your practice provides, how often is each of the following provided to people with dementia?	ProvideFam ProvideTest ProvideHome ProvideDriving ProvideFirearm ProvideHC	Continuous	Average score across all items

Table 7. Key Survey Variables Used in Nonresponse Models			
Question Text	**Variable Name(s)**	**Variable Type**	**Answer Options**
	ProvideNutrition ProvideFunction ProvideAD ProvideLegal ProvideAbuse ProvideFinance ProvideNeuro ProvideSW ProvideNamenda ProvideAmyloid ProvideBiomarker ProvideBehavior ProvideSimplify		
How much do these factors interfere with your ability to provide care for people with dementia?	InterfereTime InterfereConfidence InterfereRecords InterfereEHR InterfereLanguage InterfereFinance InterfereBilling InterfereAdmin InterfereTeam InterfereSpecial InterfereInsurance InterfereTransport InterfereCommunity InterfereResource InterfereScope	Continuous	Average score across all items

For each model in Step 1, the eligible predictors consisted of the full set of frame variables listed in **Table 6**. Frame variables that were selected in at least five of the 12 LASSO models predicting the key survey items in **Table 7** were retained for nonresponse modeling. One additional variable—whether the clinician was a sole proprietor—was selected in four models but was retained because it represents a clinician-level characteristic rather than a geographic area-level measure.

For Step 2, the retained frame variables were then included as predictors in a final logistic regression model with survey response status (respondent vs nonrespondent) as the outcome. **Table 8** presents the estimated coefficients from this response-propensity model. The model was weakly predictive of response, with a pseudo-R^2 of 0.007 and an AUC of 0.559.

Table 8. Estimated Coefficients, Standard Errors, Wald Chi-Squares, and P-Values for a Model Predicting Response to the Community Clinician Survey			
Parameter	**Value**	**Estimate[a]**	**Pr > ChiSq**
Intercept			<.0001
Clinician type (ref=Neurologist)	Primary care physician	-	0.0257
	Primary care NP	+	<.0001
	PA	+	0.0002
	Psychiatrist	+	0.1457
	Psychiatric-mental health NP	+	<.0001
Low-income patients with dementia, n (ref=0-10)	11-20	-	0.5366
	>20	-	0.6596
Patients with dementia, n (ref=1-10)	11-20	-	0.9214
	21-50	+	0.8176

Table 8. Estimated Coefficients, Standard Errors, Wald Chi-Squares, and P-Values for a Model Predicting Response to the Community Clinician Survey			
	>50	+	0.7337
Census region (ref=South)	Midwest	+	0.3757
	Northeast	-	0.0042
	West	-	0.4222
Setting (ref=Outpatient [non-residential])	Part D prescriber only	-	0.0382
	any residential	+	0.1067
Provider sex	M	+	0.5423
Patients with dementia who are black or Hispanic, n (ref=0-10)	11-20	-	0.0258
	>20	-	0.0002
Sole proprietor	Y	-	0.0929
ADI National Rank[b]		+	0.0095
% <65 without Health Insurance 2021[b, c]		-	0.0005

ADI: Area Deprivation Index

[a] The reference group for each variable is not listed in the table. Estimate values are masked for disclosure protection; sign denotes whether each estimate was positive or negative.

[b] Based on clinician practice location at the county level

[c] Determined from U.S. Census data

The predicted probabilities from this model were split into deciles. For each decile, we calculated the response rate and used the inverse as the nonresponse adjustment (Little, 1993).

The following formula was used for the nonresponse adjustment factor (see **Table 9**):

$$w_{nr,c} = \frac{1}{RR_c}$$

Where:

nr = indicates this is a nonresponse adjustment (as opposed to sampling or poststratification)

c = index of the nonresponse class (decile) for each case (listed in Table 9)

RR_c = response rate calculated for that class (decile) of the clinician

Table 9. Deciles of the Predicted Response Propensity, Response Rates, and Nonresponse Adjustment Factors				
Response Propensity Decile	**Sample Size**	**Respondents**	**Response Rate (RR_c)**	**NR adjustment factor ($\frac{1}{RR_c}$)**
1	2,489	319	0.128	7.80251
2	2,490	387	0.155	6.43411
3	2,490	440	0.177	5.65909
4	2,490	457	0.184	5.44858
5	2,490	432	0.17349	5.76389
6	2,489	469	0.18843	5.30704
7	2,490	512	0.20562	4.86328
8	2,490	492	0.19759	5.06098
9	2,490	590	0.23695	4.22034
10	2,490	601	0.24137	4.14309

3.6. Poststratification

Poststratification was used to further align the respondent sample with the known distribution of clinicians in the sampling frame and to reduce residual nonresponse bias remaining after application of selection and nonresponse adjustment weights. The following sample frame variables were used for poststratification: clinician type, setting, and the number of low-income patients with dementia (split by the median). The respondent distribution was obtained using the product of the selection weight and the

nonresponse adjustment (**Table 9**). Population proportions were derived from the sampling frame, which yielded the poststratification factors presented in **Table 10**.

Table 10: Weighted Respondent Counts, Population Counts, and Poststratification Factors

Post-stratum (g)	Clinician Type	Low-income patients[a]	Population Count (N_g)	Weighted Respondents ($\widehat{N}_g$)	PS Adjustment Factor ($w_{ps,g}$)
1	Primary care physician	below	120,153	123,548.1	0.973
2	Primary care physician	above	77,282	71,416.5	1.082
3	Primary care NP	below	119,379	116,086.6	1.028
4	Primary care NP	above	44,368	48,017.9	0.924
5	PA	below	69,452	70,630.9	0.983
6	PA	above	14,821	13,472.7	1.100
7	Psychiatrist	below	14,764	14,169.5	1.042
8	Psychiatrist	above	4,799	5,458.0	0.879
9	Psychiatric-mental health NP	below	8,400	8,795.6	0.955
10	Psychiatric-mental health NP	above	3,423	3,002.1	1.140
11	Neurologist	below	5,391	4,616.8	1.168
12	Neurologist	above	9,951	10,636.8	0.936

[a] Reflects whether clinicians were below (≤10) or above (>10) the overall median number of low-income patients with dementia

The poststratification factors in each poststratum (denoted g) are calculated using the following formula:

$$w_{ps,g} = \frac{N_g}{\widehat{N}_g},$$

Where $\widehat{N}_g$ is the following:

$$\widehat{N}_g = \sum_{i=1}^{r_g} w_{h,i} \times w_{nr,c,i}$$

Where:

r_g is the number of respondents in poststratum g

w_{hi} is the selection weight for clinician i in stratum h

$w_{nr,c,i}$ is the nonresponse adjustment for clinician i who is a member of class c

3.7. Final Weight

The final survey weight is a product of the probability of selection weight, the nonresponse adjustment, and the poststratification factor.

$$FINALWEIGHT_i = w_h \times w_{nr,c} \times w_{ps,g}$$

Where:

w_h = probability of selection weight

$w_{nr,c}$ = nonresponse adjustment

$w_{ps,g}$ = poststratification factor

Using these weights in analyses of NDWS survey data, together with appropriate variance estimation procedures described below, supports inference to the national population of clinicians represented in the sampling frame while reducing bias associated with the survey design and nonresponse. Across a small

set of representative variables, the impact of weighting and clustering on variance was small to modest, with design effects ranging from 1.1 to 1.5, suggesting minimal loss of statistical efficiency due to the weighting procedure (**Table 11**).

Table 11: Estimated Unweighted Means, Weighted Means, Design-Adjusted Standard Errors, and Design Effects for Several Respondent-Level Survey Variables

Variable	Unweighted Mean	Weighted Mean	Standard Error	Design Effect
JobBurnedOut: I feel burned out from my work.	3.5	3.52	0.03	1.2
JobHoursWeek: In a typical week, how many hours do you usually work in your principal clinical job?	39.1	39.39	0.2	1.2
JobsClinical: How many paid clinical jobs do you have?	1.6	1.66	0.05	1.5
PatientDementia: What percent of the patients on your current panel have any stage of dementia?	13.5	13.67	0.3	1.1
PatientPanel: As of today, what is the approximate size of your patient panel?	874	953	18	1.2
PracticeFT: How many years have you been practicing as a full-time years?	13.7	14.1	0.2	1.1

3.8. Sampling and Weighting References

Little, R. J. A. (1993). "Post-Stratification: A Modeler's Perspective." *Journal of the American Statistical Association*, 88(423), 1001–1012.

Little, R. J. A., & Vartivarian, S. (2005). "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology*, 31(2), 161–168.

4. Example Code for Weighted Analysis and Variance Estimation

As described above, the NDWS Community Clinician survey is based on a stratified sample of individual clinicians. Following data collection, a survey weight was developed to account for probabilities of selection, nonresponse, and poststratification. The examples below demonstrate how to incorporate the final survey weight and stratification variables in analyses of Community Clinician survey data using SAS or Stata.

To obtain correct standard errors and confidence intervals in statistical analyses, the sample design, including 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 and survey weights specified via the WEIGHT statement. For the Community Clinician Survey, the stratification variable is `stratum` and the survey weight is `finalweight`.

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 design variables, while the PUF omits the stratification variable (stratum) for disclosure protection. Although the stratification variable is not available in the PUF, valid weighted point estimates can still be produced using the final survey weights alone. This will yield unbiased point estimates, however, omitting stratification information will result in slightly larger variance estimates.

- ***Producing a weighted estimate:***
  The following SAS code will generate a weighted estimate for `PracticeFT`, a continuous variable, incorporating sample design elements:

  The following code uses Community Clinician data to provide two estimates for the mean (green boxes) of `PatientPanel`: the first uses the PUF data set (i.e., without the `strata` element) and the second uses LINKAGE data and incorporates the sample design features (right table column). The only difference between the PUF and RUF means are in the respective standard errors (red boxes); those generated from the RUF data, which includes stratum, are slightly smaller (18.82 [PUF] vs. 18.25 [RUF]).

Public Use File (PUF)

```
PROC SURVEYMEANS DATA= cc_puf;
  * STRATA stratum;
  WEIGHT FinalWeight;
  VAR PatientPanel;
RUN;
```

N	Mean	Std Error of Mean	95% CL for Mean	
4617	952.532142	18.821723	915.632567	989.431716

Restricted Use File (RUF)

```
PROC SURVEYMEANS DATA= cc_ruf;
  STRATA stratum;
  WEIGHT FinalWeight;
  VAR PatientPanel;
RUN;
```

N	Mean	Std Error of Mean	95% CL for Mean	
4617	952.532142	18.252420	916.748638	988.315646

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 Community Clinician survey, this includes the weight and stratum. In this case, weight is declared via `[pweight=FinalWeight]` and stratum as `strata(stratum)`.

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

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

```
svy: mean PracticeFT
```