

NDWS Weighting Tutorial video transcript, Sept. 28, 2026

Welcome! In this NDWS tutorial, we will explain why survey weights matter, which variables to declare in the survey design, and provide code examples in SAS, Stata, and R.

Why Use Weights?

So, why do weights matter? A regular frequency table is useful because it tells us how many respondents are in the file, but when we want an estimate for the population represented by the survey, we use the final weight and the survey design. The weight adjusts the contribution of each respondent, and the design statement helps the software produce appropriate standard errors and confidence intervals.

Design Declaration: Three Ingredients

The design declaration usually needs three pieces of information. The final weight, the stratification variable, and the cluster or PSU variable. The exact weighting variable can differ by release of the NDWS file, so the safest approach is to check the variable name in your file or in the NDWS User Guide.

SAS Demonstration: PROC SURVEYMEANS

For SAS, the example provides the weighted mean of the variable JobOverall, which represents the overall level of satisfaction with the job.

The DATA equals line right after the PROC SURVEYMEANS command points to the analytic file. Here, we used the NDWS Assisted Living Staff Survey Restricted Use file, AL_staff_RUF.

The STRATA line identifies the stratum variable, and CLUSTER identifies the PSU or cluster variable. NDWS surveys have Restricted Use Files, RUFs, and Public Use Files, PUFs. In this example, we used RUF. If you use PUF, the cluster variable will be pSampleCluster instead of NDWS_FACID.

The WEIGHT line identifies the final weight, and VAR identifies the outcome we want to examine. Here, the outcome is the variable JobOverall.

Stata Demonstration

In Stata, the workflow is two steps. First, declare the survey design using the command svyset. If you have the cluster variable, put it before the weight statement. For NDWS Community Clinician surveys, the sampling frame is a list of all eligible clinicians without clusters.

So, in the code example, we left the cluster variable blank before the weight statement. Then include the final weight as pweight and stratum as the stratification variable. For the Community Clinician PUF, omit strata when the file does not provide a stratum variable.

Second, run analyses with the `svy` prefix. Our example here gives the weighted mean and standard error of the variable `PatientPanel`, which is the size of the clinician's patient panel. One thing to note is that Stata is case sensitive, so be careful with uppercase and lowercase.

R Demonstration: Survey Package

In R, the survey package is really helpful. The `svydesign` line creates the survey design object. Then, in the next line, `ids` is the PSU or cluster variable; `strata` is stratum; `weights` is `FinalWeight`; `data` is the NDWS data frame; and `nest=TRUE` is commonly used when PSUs are nested within strata. Then the analysis functions are applied to the design object.

Our example uses the NDWS Nursing Home Staff Survey PUF. If you use the Community Clinician files, which do not contain a cluster variable, use `ids = ~1`. R is also case sensitive, so watch uppercase and lowercase.

Takeaways

The code examples might sound complicated, but the closing message is simple: use unweighted counts to understand the respondent's sample and to check missingness.

Use the survey design for population estimates. In reports and presentations, clearly label the number: is it an unweighted respondent count, a weighted population total, or a weighted percentage? That prevents the most common interpretation mistake in weighted survey analysis.

Differences in Each Language

This table translates the same design components across the three software environments. `FinalWeight` is specified with `WEIGHT` in SAS, `pweight` in Stata, and `weights` in R. The stratum and cluster variables are likewise supplied through `STRATA` and `CLUSTER` in SAS, `strata()` and the PSU position in Stata's `svyset` command, and `strata` and `ids` in R.

When there is no cluster variable, omit that component in SAS or in Stata, but in R, use `ids = ~1`. SAS and Stata handle clusters nested within strata through the combined design specification, whereas R uses `nest=TRUE`. SAS `PROC SURVEYMEANS` and Stata `svy:mean` report 95% confidence intervals, while R requires a separate call, `confint()`. If a file does not contain a stratum variable, omit that component. In Nursing Home and Assisted Living PUFs, use `pSampleCluster` as the cluster or PSU variable.

NDWS Website

We hope this video gets you excited to explore the NDWS dataset. We have detailed documentation on our survey design and all the good stuff you'd like to see on our website.

Acknowledgement

Thank you for watching this video. Please visit ndws.org for the most up-to-date information and see our special collection in the Journal of the American Geriatrics Society.

This video was produced by Rona Hu, Esther Friedman, James Wagner, Martha Pennington, and Donovan Maust, with funding from the National Institute on Aging. Comments and questions may be sent to info@ndws.org. Thanks for listening and for your interest.