

Uses of State Information in NHATS Analyses

October 12, 2026

The National Health and Aging Trends Study (NHATS) is designed to produce nationally representative estimates of Medicare beneficiaries aged 65 years and older. Although state identifiers are available in the restricted-use data, NHATS was not designed to support representative estimates for individual states. In particular, the survey weights are not calibrated to state population totals, and sample sizes within individual states may be insufficient for reliable state-level estimation. Consequently, **the use of state information should be limited to analyses that are consistent with the survey design and disclosure protection requirements.**

General rules that apply to all analyses that use state identifiers

State characteristics should **not** be presented in a way that allows individual states where NHATS respondents live to be identified. State-specific output may not be exported, reported, or included in any documentation or publications.

Information intended for export or publication outside of the enclave may not contain:

- Individual state names, including the states that NHATS respondents are not observed
- Counts or tabulations of states in the sample
- State-specific NHATS sample sizes
- State-group level range including maximum and minimum values
- State-specific random effects, predicted values, residuals, or coefficients
- Rankings, maps, plots with state-level value, or other displays comparing states
- Identification of states as unusually high, low, best, worst, or outliers
- Any other output not specified above that could be used to infer the identification or characteristics of an individual state

Below are restriction levels and guidelines for specific state-level analyses that are often requested to be conducted with NHATS data, starting with those that are not feasible or permitted.

Not feasible / Not permitted

1. Producing State-Level Estimates and Making Inferences to Individual States

Researchers should not use NHATS to estimate state-specific prevalence, population totals, or other descriptive statistics that are intended to represent the population of an individual state.

Examples of analyses that are not permitted:

- Estimating the prevalence of dementia in a given state.
- Estimating the number of older adults receiving unpaid care in a given state.
- Ranking states according to dementia prevalence.

Permitted with Restrictions

1. Grouping States According to Shared Characteristics

Researchers may classify states into broader groups based on externally defined characteristics and compare individuals living in one group of states with individuals living in another group, as long as there are at least 6 states per group.

Possible grouping variables include:

- States with and without certain policies (e.g., paid family leave)
- States with a certain level on contextual variable (e.g., Medicaid generosity)

Additional Rules Apply to Applications for and Approved Projects Grouping States According to Shared Characteristics

- For applications that wish to use state-groups, between 7 and 43 states must be in each group, in the population.
- The count of states underlying the NHATS sample must meet a minimum threshold of 6 states in each group to be exported from the enclave. For analyses that group states by more than one variable (for example, by level of Medicaid generosity and year) each resulting subgroup must include at least 6 states.
 - o Researchers should note that it is recommended to have at least 20 counties per group to compute variance and reduce estimation issues.

- Counts of state-groups and individuals in state-groups must be tabulated for vetting and included in supporting documentation but are not releasable from the enclave and should not be included in publications.
- Each group or subgroup (i.e., cells in cross-tabs) involving grouped states must contain > 10 individuals.

2. Using the State Identifier Directly in Statistical Models

The state identifier may be used as a fixed-effect variable or a random-effect variable in statistical models, such as linear or logistic regression, fixed effects models, multilevel random-effect models. However, because NHATS was not designed to support inference for individual states and state-specific outputs may create disclosure risks, coefficients for state-level parameters may not be removed from the enclave.

Permitted outputs include:

- Overall state-level variance components
- Intraclass correlation coefficients
- Model-fit statistics

Feasible and Permitted

1. Controlling for Census Divisions

Controlling for Census Division (e.g. West North Central, Middle Atlantic) or Census Region (e.g., Northeast, Midwest, South, and West) is allowed because, although each Division/Region may not have more than 6 states, it combines multiple states that contains sufficient county and individual sample sizes for survey estimation. Cell thresholds of > 10 individuals still apply.

2. Linking Respondents to State-Level Continuous Contextual Characteristics

Researchers may merge external state-level contextual information to NHATS respondents in the enclave and evaluate whether those contextual characteristics are associated with individual outcomes. Examples of contextual variables include Medicaid generosity, state unemployment rate or cost-of-living index. (High and low values are not permitted unless > 10 individuals have those values.)