

Deriving Weighted Estimates and Calculating Standard Errors for States and Subpopulations for the NAM CAHPS Survey

The Centers for Medicare and Medicaid Services' (CMS) Center for Medicaid and CHIP Services (CMCS) contracted with NORC at the University of Chicago and its partner, Thoroughbred Research Group, to collect the first-of-its-kind nationwide adult Medicaid (NAM) Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey. The 2014-2015 NAM CAHPS survey provides baseline nationwide and state-level estimates of adult Medicaid enrollees' experience of care, including access to and utilization of services, across different financing and delivery models and population groups.ⁱ This brief provides an introduction to the survey weights developed for the 2014-2015 NAM CAHPS respondent-level data,ⁱⁱ as well as instructions for researchers with permission to access the microdata to derive estimates and calculate standard errors for states and subpopulations using the survey weights.ⁱⁱⁱ

SUMMARY OF SAMPLE DESIGN AND WEIGHTING UTILIZED FOR THE 2014-2015 NAM CAHPS SURVEY

Sample Design

The target population for the 2014-2015 NAM CAHPS survey was adults ages 18 and older as of December 31, 2013, who were enrolled in Medicaid for each month of the first quarter of federal fiscal year (FFY) 2014 (FFY 2014 Q1, October 2013 – December 2013), who were not residing in an institutional setting, and were not part of a Family Planning Waiver.

The sampling frame for the 2014-2015 NAM CAHPS survey was designed to capture four key subgroups of adult Medicaid enrollees. The main criteria for stratifying the sampling frame were state (including the District of Columbia) and the following four mutually exclusive enrollee groups based on program eligibility:

- adults dually eligible for Medicare and Medicaid (Full Duals);
- adults (non-duals) with disabilities (Persons with Disabilities);
- adults (non-duals, non-disabled) enrolled in a managed care organization (Managed Care, or MC); and,
- adults (non-duals, non-disabled) who obtained care from a fee-for-service provider or were enrolled in a primary care case management plan (FFS-PCCM).

The sampling frame was constructed from each state's FFY 2014 Q1 Medicaid Statistical Information System (MSIS) eligibility file, where available. For ten states that did not have these data available in MSIS, state Medicaid administrators provided a standardized data extract from each state's internal eligibility systems to create a comparable sampling frame, following CMCS/NORC-developed specifications. A systematic sample was selected within each of the four strata, within each state. The total sample was composed of over 1.2 million cases, averaging approximately 29,000 sampled enrollees per state.

Weights

The sample was drawn and the survey weights were constructed such that the 2014-2015 NAM CAHPS survey respondents were representative of the adult Medicaid population in the states. The survey weight (FinalWeight) was derived to adjust for variation in sampling rates and differential response rates to produce robust statistical estimates at the state level. The weighting scheme involved three steps:

- Base sampling weights;

- Adjustment for ineligible adults; and
- Adjustment for nonresponse among eligible adults.

For more details about the sample design and survey weighting procedures, please see the [2014-2015 NAM CAHPS Methodology Report](#).

USE OF SURVEY WEIGHT VARIABLES

It is recommended that NAM CAHPS survey data users weight their results when producing unbiased estimates of population characteristics. All estimates provided in the summary tables, charts, and maps available through the NAM CAHPS Data Resource Website have been weighted to one of the four Medicaid enrollee strata within the state. Researchers with permission to access the microdata through the 2014-2015 NAM CAHPS Data Repository should use the following variables to properly weight their results:

- FINALWEIGHT- Provides a weight for each survey respondent who participated in the 2014-2015 NAM CAHPS survey.
- STRATUM- Provides the sample stratum.
- STATE_STRATUM- Provides the sample stratum within each state. This variable should be utilized when asked about a stratum-level variable. The variable, STRATUM, is not sufficient by itself for accurate calculation of weighted estimates and associated standard errors.

The survey weights must be used to ensure that estimates derived from NAM CAHPS survey respondents approximate the population estimates for adult Medicaid enrollees in the state, sample stratum, or other subpopulation of interest.

KEY VARIABLES AVAILABLE IN THE 2014-2015 NAM CAHPS SURVEY DATA

A list of the key variables that are available to researchers for state and subpopulation estimates using the 2014-2015 NAM CAHPS survey data available through the NAM CAHPS Data Resource Website is provided in Table 1.

Table 1. Variables for State and Subpopulation Estimates in the NAM CAHPS Data Resource Website

Variable Name	Variable Label	Variable Values
Sample stratum	Sample stratum	Full Duals Persons with Disabilities Managed Care (MC) Fee-for-service or Primary care case management plan (FFS-PCCM)
Age	Age of respondent	18 to 34* 35 to 44 45 to 54 55 to 64 65 or older*
Education	Education level of respondent	8th grade or less Some high school, but did not graduate High school graduate or GED More than high school degree*
Sex	Sex of respondent	Male Female
Ethnicity [^]	Ethnicity of respondent	Not Hispanic/Latino origin Hispanic, Latino/a, or Spanish origin
Mode	Mode of survey completion	Phone Mail
Race [~]	Race of respondent	White Black/African American Asian* Other Race*
Race/ethnicity [^]	Race/ethnicity of respondent	White, Not Hispanic/Latino origin Black/African American, Not Hispanic/Latino origin Asian, Not Hispanic/Latino origin* Other Race, Not Hispanic/Latino origin* Hispanic, Latino/a, or Spanish origin

*The variables age, education, race, and race/ethnicity were recoded to collapse survey response categories in the summary tables, charts, and maps available through the NAM CAHPS Data Resource Website to minimize the risk of disclosing identifiable information. Please see the footnotes available beneath each frequency table, chart, and map for details about the recodes in specific tables.

[^]Ethnicity is coded as "Hispanic, Latino/a, or Spanish Origin" if at least one category of Hispanic/Latino origin was selected (from items Q68a, Q68b, Q68c, Q68d, Q68e). Ethnicity is coded as "Not Hispanic/Latino origin" if the respondent selected "No" for Q68a.

^Race is recoded as follows to facilitate analysis and minimize the risk of disclosing respondent privacy: all Asian categories (Q69d, Q69e, Q69f, Q69g, Q69h, Q69i, Q69j) are collapsed into one combined "Asian" category and all races other than White (Q69a) and Black/African American (Q69b) are collapsed into one combined "Other race" category. The "Other race" category includes cases where multiple race categories were selected and the following races: American Indian or Alaska Native (Q69c), Native Hawaiian (Q69k), Guamanian or Chamorro (Q69l), Samoan (Q69m), Pacific Islander (Q69n) and some other race (Q69o).

>Race/ethnicity is recoded as follows to facilitate analysis and minimize the risk of disclosing identifiable information: cases where at least one category of Hispanic/Latino origin was selected (from items Q68a, Q68b, Q68c, Q68d, Q68e) are recoded as "Hispanic, Latino/a, or Spanish Origin", all Asian categories (Q69d, Q69e, Q69f, Q69g, Q69h, Q69i, Q69j) are collapsed into one combined "Asian" category and all races other than White (Q69a) and Black/African American (Q69b) are collapsed into one combined "Other race" category. The "Other race" category includes cases where multiple race categories were selected and the following races: American Indian or Alaska Native (Q69c), Native Hawaiian (Q69k), Guamanian or Chamorro (Q69l), Samoan (Q69m), Pacific Islander (Q69n) and some other race (Q69o). Race/ethnicity includes both individuals who indicated they were not of Hispanic, Latino/a, or Spanish Origin and those whose ethnicity could not be ascertained because of missing data, if a non-missing response was provided for race.

While these enrollee characteristics can be used to identify subpopulations within the summary tables, charts, and maps available through the NAM CAHPS Data Resource Website, additional subpopulations can be identified through the use of responses to other survey questions available in the complete respondent-level microdata. It is important that researchers with access to the microdata use appropriate survey weights when deriving estimates and also derive robust standard errors to approximate the reliability of each estimate regardless of the population of interest.

CALCULATING STANDARD ERRORS AND DRAWING INFERENCES

The standard error is a measure of the statistical accuracy of an estimate. As described above, the 2014-2015 NAM CAHPS survey used a stratified, single-stage sampling design, where the stratifiers were state and stratum within state (Full Duals, Persons with Disabilities, Managed Care, and FFS-PCCM). The standard error should be calculated using the Taylor series linearization method, which takes into account the complex sample design.^{iv}

A more straightforward way to interpret what a standard error means is to use it to calculate a 95 percent confidence interval. The 95 percent confidence interval (CI) for a proportion (P) can be attained by multiplying the standard error (SE) by 1.96.

Formula: 95% CI = $P \pm 1.96 \times SE$

In using the data, researchers may also want to calculate other statistics that can be used to assess the reliability of data, such as the margin of error and coefficient of variation. For more details about variance estimation, please see the [2014-2015 NAM CAHPS Methodology Report](#).

The final section of this report provides sample code for common statistical packages to help calculate weighted estimates and standard errors.

SAMPLE CODE FOR COMMON STATISTICAL PACKAGES

The sample code below shows how a data user with access to the 2014-2015 NAM CAHPS microdata could generate a cross-tabulation of weighted data and calculate standard errors using Taylor series linearization using different statistical packages.

SAS

```
proc surveyfreq data=DATANAME;
table VARIABLE1*VARIABLE2 / row col ;
strata STATE_STRATUM;
weight FINALWEIGHT;
ods output crosstabs=TABLENAME;
run;
```

STATA

```
svyset [pweight= FINALWEIGHT], strata(STATE_STRATUM)
svy: tab VARIABLE1 VARIABLE2
```

R

```
mydesign<-svydesign(weights=~ FINALWEIGHT, strata=~ STATE_STRATUM ,data=mydata)
svytable(~VARIABLE1+VARIABLE2, design=mydesign)
```

About the NAM CAHPS Survey

In 2013, the Centers for Medicare & Medicaid Services' (CMS) Center for Medicaid and CHIP Services (CMCS) contracted NORC at the University of Chicago (NORC) and its partner, Thoroughbred Research Group to collect the first-ever nationwide adult Medicaid (NAM) Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey. The goal of the 2014-2015 NAM CAHPS survey is to obtain nationwide and state-by-state estimates of adult Medicaid enrollees' experience of care, including access and utilization of services, across different financing and delivery models (e.g., managed care and fee-for-service) and population groups (e.g., individuals who are dually eligible for Medicare and Medicaid and individuals with disabilities who are not dually eligible).

The sample for the 2014-2015 NAM CAHPS survey was designed to capture four key subgroups of adult Medicaid enrollees. The main criteria for stratifying the sampling frame were state (including the District of Columbia) and the following four mutually exclusive enrollee groups based on program eligibility:

- adults dually eligible for Medicaid and Medicare (Full Duals);
- adults (non-duals) with disabilities (Persons with Disabilities);
- adults (non-duals, non-disabled) enrolled in a managed care organization (Managed Care, or MC); and,
- adults (non-duals, non-disabled) who obtained care from a fee-for-service provider or were enrolled in a primary care case management plan (FFS-PCCM).

Forty-six states and the District of Columbia participated in the 2014-2015 NAM CAHPS survey. The target population for the survey was adults ages 18 and older as of December 31, 2013, who were enrolled in Medicaid during first quarter of federal fiscal year (FFY) 2014 (FFY 2014 Q1, October 2013 – December 2013), and who were not residing in an institutional setting, and were not part of a Family Planning Waiver. Data collection was conducted from December 14, 2014 through July 27, 2015, across four waves. The questionnaire was administered first through mail, and then with telephone follow-up where necessary, and was available in both English and Spanish. This effort resulted in 272,679 enrollees completing the survey, with an overall response rate of 23.6%.

Learn more about the NAM CAHPS survey and explore this data at nam-cahps.medicaid.gov.

REFERENCES

ⁱ Alaska, New Hampshire, North Dakota, and Wisconsin did not participate in the 2014-2015 NAM CAHPS survey.

ⁱⁱ Two types of respondent level data will be available for the 2014-2015 NAM CAHPS: 1) a Public Use File (PUF) of fully de-identified respondent-level survey data, and 2) a Limited Data Set (LDS) of detailed respondent-level survey data. Visit the [Use the Data](#) page of the NAM CAHPS data resource website to learn more about how to access NAM CAHPS survey data.

ⁱⁱⁱ For more details about the 2014-2015 NAM CAHPS Data Repository files please see the [2014-2015 NAM CAHPS Codebook](#).

^{iv} Wolter, K.M. (2007). Introduction to variance estimation (2nd ed.). New York: Springer-Verlag.