National Center for Health Statistics Research Data Center
National Ambulatory Medical Care Survey Health Center Component
2023 Restricted Use File Data Dictionary

This document contains the data dictionary for the 2023 National Ambulatory Medical Care Survey (NAMCS) Health Center (HC) Component data available at the National Center for Health Statistics (NCHS) and Federal Research Data Centers (RDC). The NAMCS HC Component collects data on patient care in federally qualified health centers (FQHCs) to describe patterns of health care delivery and utilization in the United States. The 2021 survey year was the first in which NAMCS collected FQHC data from electronic health record (EHR) systems, which continued for the 2023 survey year.

Participating health centers were asked to submit electronic health record (EHR) data for all encounters in calendar year 2023 in accordance with the standard and format requested for NAMCS, which is the Health Level Seven International (HL7) Clinical Document Architecture (CDA®) R2 Implementation Guide (IG): National Health Care Surveys, DTSU Release 1, Release 1.2, or Release 3-US Realm, from here on referred to as the IG. The IG was created by NCHS for use by the National Health Care Surveys. Some FQHCs were unable to provide data due to limitations with their EHR systems and therefore submitted custom extracts using the data elements and formats in the IG as a template. These data included personal patient identifiers such as name, address, and social security number when it is available; date of visit; diagnoses and services provided or ordered during the visit; reason for visit; and clinical notes. A full listing of possible data elements and processing specifications can be found here: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=385. Despite the request for data in a standard format, some records received by NCHS still included values for certain variables that did not conform with the IG. In these instances, nonconforming values for certain variables were recoded by NCHS to align with the IG, as described in the tables below.

In calendar year 2023, 95 of 315 FQHCs submitted data on 9,012,885 visits from January 1, 2023 to December 31, 2023, for an unweighted response rate of 30.16% and a weighted response rate of 26.96%. Response rates were calculated using guidance from the American Association for Public Opinion Research's publication of "Response Rates – An Overview", available here: https://aapor.org/publications-resources/education-resources/response-rates/. Specifically, response rates for the NAMCS HC Component were calculated using Response Rate 4 in the AAPOR Response Rate Calculator 5.1.

The 2023 NAMCS HC Component data are weighted and can be used to produce nationally representative estimates of visits at FQHCs. NCHS will release a public use file for the 2023 NAMCS HC, and NCHS is making a more restricted set of data available for users to request in the RDC. Users should pay close attention to variance estimates and NCHS presentation standards for proportions (found here: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf) and for counts and rates (found here:

https://www.cdc.gov/nchs/data/series/sr_02/sr02-200.pdf) before publishing any weighted estimates. Examples of SAS-callable SUDAAN, Stata, and R code are provided at the end of this document to show how the weight and design variables should be used to produce weighted estimates.

Importantly, data users must review and consider the analytic guidelines provided at the end of this document. Some health centers were unable to provide certain data elements, and therefore users must account for this missingness by conducting complete case analysis and normalizing weights.

The data are made available in relational tables, and the names and descriptions of the data tables are listed below.

Table Name	Description	Observation Unit
Visit	Visit date, health center ID, and pregnancy flag	Visit
Patient	Patient demographics	Visit
Condition	Conditions and diagnoses associated with the visit	Condition or diagnosis
Procedure	Procedure and service codes associated with the visit	Procedure or service
Vital	Biometrics and vitals recorded at the visit	Biometric or vital
Medication	Medications recorded at the visit	Medication
Lab	Lab tests that were conducted or recorded at the visit	Lab
Weight	Visit weights and design variables to produce national estimates	Visit

The table below shows the join relationships between the tables.

- 1→ 1 indicates one-to-one visit links.
- 1→ ∞ indicates one-to-many visit to observation links.
- ∞ → ∞ indicates many-to-many observation to observation links.

Table Name	Visit	Patient	Condition	Procedure	Vital	Medication	Lab	Weight
Visit		1→ 1	1→ ∞	1→ ∞	1→ ∞	1→ ∞	1→ ∞	1→ 1
Patient	1→ 1		1→ ∞	1→ ∞	1→ ∞	1→ ∞	1→ ∞	1→ 1
Condition	∞ → 1	∞ → 1		∞ → ∞	∞ → ∞	∞ → ∞	∞ → ∞	∞ → 1
Procedure	∞ → 1	∞ → 1	∞ → ∞		∞ → ∞	∞ → ∞	∞ → ∞	∞ → 1
Vital	∞ → 1	∞ → 1	∞ → ∞	∞ → ∞		∞ → ∞	∞ → ∞	∞ → 1
Medication	∞ → 1	∞ → 1	∞ → ∞	∞ → ∞	∞ → ∞		∞ → ∞	∞ → 1
Lab	∞ → 1	∞ → 1	∞ → ∞	∞ → ∞	∞ → ∞	∞ → ∞		∞ → 1
Weight	1→ 1	1→ 1	1→ ∞	1→ ∞	1→ ∞	1→ ∞	1→ ∞	

Contents

Visit Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
SOURCE	Data Source	Character	"IG" "Custom Extract"	All visits are pulled from EHRs according to the IG, however, as mentioned in the introduction, some health centers needed to produce custom extracts of their records to conform with the format needed for processing via the IG. These health center records are coded as "Custom Extract" instead of "IG".
HC_ID	NCHS-assigned health center identifier	Character		
VISIT_START_DATETIME	Visit start date	DateTime		Values are in DATETIME22.3 format, for example "01JAN2023 : 09: 00 : 00.000".
PREGNANT	Indicator that the visit is by a pregnant female, regardless of the reason for visit	Character	"Pregnant" ""= Missing	Derived according to the IG, which identifies a visit by a pregnant person using the SNOMED CT code 77386006.

Patient Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
PATIENT_ID	NCHS-assigned patient identifier	Numeric		Included to allow users to assess demographic values across visits by the same patient. Not to be used for patient-level estimates.
PATIENT_AGE_D	Patient age in years, months, or days.	Numeric		The values in this variable represent either days, months, or years. Therefore, the variable PATIENT_AGE_UNIT_D must be used with this variable. Please note, a small number of visits have negative patient age values. These have been left as extracted from EHRs, and users will need to decide how to handle outlier values.
PATIENT_AGE_UNIT_D	Unit for patient age	Character	1 = Years 2 = Months 3 = Days	
PATIENT_SEX_D	Patient sex	Character	-9 = Missing/Unknown 1 = Male 2 = Female	
MARITAL_STATUS_R	Recoded patient marital status	Character	-9 = Missing/Unknown A = Annulled C = Common Law D = Divorced I = Interlocutory L = Legally Separated M = Married P = Polygamous S = Never Married T = Domestic Partner U = Unmarried W = Widowed O = Other	Derived according to the IG, with non-conforming values recoded by NCHS to align with values in the IG (for example, “Married” recoded to “M”).

RDC Variable Name	Description	Format	Value	Notes
PATIENT_RACE_R	Recoded patient race	Character	-9 = Missing/Unknown 1 = American Indian or Alaska Native 2 = Asian 3 = Black or African American 4 = Native Hawaiian or Other Pacific Islander 5 = White 6 = Other race	Derived according to the IG, with non-conforming values recoded by NCHS to align with values in the IG (for example, "African American" recoded to 3).
PATIENT_ETHNICITY_R	Recoded patient Hispanic ethnicity	Character	-9 = Missing/Unknown 1 = Hispanic or Latino 2 = Not Hispanic or Latino	Derived according to the IG, with non-conforming values recoded by NCHS to align with values in the IG (for example, "Hispanic or Latino" recoded to 1).

Condition Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
CONDITION_CODE	Original diagnosis code extracted from EHR record	Character		This is included so that users can see original diagnosis values prior to being recoded to ICD-10-CM. We recommend using the recoded variable below.
CONDITION_CODESYS_NAME	Code system of original diagnosis code (CONDITION_CODE)	Character	"ICD-10" "ICD-9-CM" "ICD-10-CM" "SNOMED CT" "" = Missing	
CONDITION_CODE_R	Recoded diagnosis code into ICD-10-CM format	Character		CONDITION_CODE was recoded to CONDITION_CODE_R, with all values standardized to ICD-10-CM coding. We recommend using this recoded variable when assessing diagnoses at health center visits. In this variable, periods have been removed from the 4th position of the value; for instance, "F17.200" (Nicotine dependence, unspecified, uncomplicated) is represented as "F17200". CONDITION_CODE_R is missing from 22.7% of condition records, and about 70.0% of those records have a CONDITION_CODE present that could not be recoded to ICD-10-CM. Users can assess the original CONDITION_CODE in those instances to see if there is useful information.
CONDITION_CODESYS_NAME_R	Code system of recoded diagnosis code (CONDITION_CODE_R)	Character	"ICD10-CM" "" = Missing	All condition codes were recoded to ICD-10-CM.

RDC Variable Name	Description	Format	Value	Notes
DIAGNOSIS_TYPE	Indicator for whether a condition code is primary or not	Character	"Primary" "Diagnosis" "" = Missing	"Primary" diagnosis is identified by the presence of LOINC code 52534-5, as specified in the IG. Diagnosis type is missing in 59.4% of condition records, so DIAGNOSIS_TYPE may not be particularly useful.
CONDITION_STATUS	Indicator for whether a condition code is active or not	Character	"Active" "Inactive" "" = Missing	"Active" versus "Inactive" conditions are identified by SNOMED CT codes of 55561003 and 73425007, respectively, as specified in the IG. CONDITION_STATUS is missing in 37.1% of condition records.

Procedure Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
PROCEDURE_CODE_R	Recoded procedure code	Character		This variable was recoded from PROCEDURE_CODE to remove values that were determined to not meet procedure code formats, a majority of which had values of “null”, “NULL”, or “OTH” for PROCEDURE_CODE.
PROCEDURE_CODESYS_NAME_R	Recoded procedure code system name	Character	“CPT” “LOINC” “SNOMED” “” = Missing	

Vital Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
VITAL_TYPE_R	The type of biometric or vital recorded at the visit	Character		Not all visits have a vital record. Users will need to use VITAL_TYPE_R, VITAL_UNIT_R, and VITAL_VALUE_R together when assessing vitals at health center visits. Not all vital records have values for all three variables.
VITAL_UNIT_R	Unit of the biometric or vital	Character		
VITAL_VALUE_R	Value of the biometric or vital	Numeric		NOTE: NCHS removed vital values that were non-numeric, but did not alter any numeric vital values. Users should review this variable for any outliers within the vital type and vital unit of interest before analysis.

Medication Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
MEDICATION_CODE_R	Recoded medication code	Character		Medication codes are either RxNorm or National Drug Codes (NDC), with the large majority of codes as RxNorm. This variable pairs with MEDICATION_CODESYS_NAME_R. MEDICATION_CODE_R is missing in 21.5% of medication records because they were originally missing or could not be recoded.
MEDICATION_CODESYS_NAME_R	Recoded code system of the medication code value	Character	"RxNorm" "NDC" "" = Missing	This variable was recoded to create consistent values.
MEDICATION_CODE_NLP	Recoded and recovered[1] medication code	Character		This variable pairs with MEDICATION_CODESYS_NAME_NLP. MEDICATION_CODE_NLP is missing in 16.9% of medication records because they do not have a recoded value and could not be matched using Natural Language Processing (NLP).
MEDICATION_CODESYS_NAME_NLP	Code system of the recoded and recovered medication code values	Character	"RxNorm" "" = Missing	All recoded and recovered medication codes were provided in RxNorm.
MEDICATION_STATUS	Indicator for whether a medication is active, completed, held or suspended	Character	"active" "completed" "held" "suspended" "" =Missing	MEDICATION_STATUS is missing for 1.2% of records where MEDICATION_CODE_NLP is non-missing and 1.6% of records where MEDICATION_CODE_R is non-missing.

[1] Some records missing a medication code were recovered using the medication name and an NLP matching technique to populate the RxNorm medication code. The NLP technique utilized Jaro similarity fuzzy matching to recover approximately 2.2 million records (out of the 4.6 million that were attempted to match) that are either missing a medication code or do not have an appropriate code system name, but have a corresponding medication name. This technique focused on matching the provided medication name, absent of the medication dosage and route, with an RxNorm code from a standardized library. An exact match was produced for 85.5% of the 2.2 million records that were recovered. The remaining 14.5% of recovered records were populated with a correct RxNorm code approximately 80% of the time based on an assessment of a random sample of non-exact matches. This variable incorporates both the recoded medication code and all recovered missing and NDC code system medication records.

Lab Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
LAB_TESTCODE_R	Recoded lab test code	Character		All recoded lab test codes are LOINC codes.
LAB_QUANT_RESULT_30	Recoded lab code	Character		This variable is only available for the top 30 lab tests in the database, which are provided in the note below the table. The lab result includes the lab test value and oftentimes the unit for the lab test. NCHS did minimal cleaning to the results in this variable outside of ensuring no sensitive information was contained in values provided. NOTE: not all records that have a LOINC code in LAB_TESTCODE_R have a value in LAB_QUANT_RESULT_30.

NOTE: LAB_QUANT_RESULT_30 is only available for the 30 most common LAB_TESTCODE_R values in the 2023 database. These top 30 LOINC codes and their descriptions are:

1. 2160-0: Creatinine [Mass/volume] in Serum or Plasma
2. 2951-2: Sodium [Moles/volume] in Serum or Plasma
3. 2823-3: Potassium [Moles/volume] in Serum or Plasma
4. 2075-0: Chloride [Moles/volume] in Serum or Plasma
5. 17861-6: Calcium [Mass/volume] in Serum or Plasma
6. 3094-0: Urea nitrogen [Mass/volume] in Serum or Plasma
7. 2345-7: Glucose [Mass/volume] in Serum or Plasma
8. 2028-9: Carbon dioxide, total [Moles/volume] in Serum or Plasma
9. 3097-3: Urea nitrogen/Creatinine [Mass Ratio] in Serum or Plasma
10. 718-7: Hemoglobin [Mass/volume] in Blood
11. 4544-3: Hematocrit [Volume Fraction] of Blood by Automated count
12. 785-6: MCH [Entitic mass] by Automated count
13. 787-2: MCV [Entitic volume] by Automated count
14. 789-8: Erythrocytes [#/volume] in Blood by Automated count
15. 6690-2: Leukocytes [#/volume] in Blood by Automated count

16. 786-4: MCHC [Mass/volume] by Automated count
17. 788-0: Erythrocyte distribution width [Ratio] by Automated count
18. 777-3: Platelets [#/volume] in Blood by Automated count
19. 1975-2: Bilirubin total [Mass/volume] in Serum or Plasma
20. 6768-6: Alkaline phosphatase [Enzymatic activity/volume] in Serum or Plasma
21. 2885-2: Protein [Mass/volume] in Serum or Plasma
22. 1920-8: Aspartate aminotransferase [Enzymatic activity/volume] in Serum or Plasma
23. 1751-7: Albumin [Mass/volume] in Serum or Plasma
24. 1742-6: Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma
25. 1759-0: Albumin/Globulin [Mass Ratio] in Serum or Plasma
26. 10834-0: Globulin [Mass/volume] in Serum by calculation
27. 4548-4: Hemoglobin A1c/Hemoglobin total in Blood
28. 706-2: Basophils/100 leukocytes in Blood by Automated count
29. 736-9: Lymphocytes/100 leukocytes in Blood by Automated count
30. 713-8: Eosinophils/100 leukocytes in Blood by Automated count

Weight Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
HC_ID	NCHS-assigned health center identifier	Numeric		
POPHC	Estimated total number of in scope health centers in the sampling stratum from which the health center was selected	Numeric		
STRATUM_S	Scrambled sampling stratum from which the health center was selected	Numeric		Original sampling stratum values were scrambled by NCHS.
TOTVISWT	Visit weight	Numeric		

Coding Examples for Weighted Estimates

Below are examples in SAS-callable SUDAAN, Stata, and R of how to use weights and design variables for producing visit-level weighted estimates of and for approximating variance.

Software	Visit-level estimates
SAS-callable SUDAAN	PROC {procedure} DATA = {input data set} DESIGN = WOR {STATISTIC TYPE}; NEST STRATUM_S HC_ID / MISSUNIT; TOTCNT POPHC _ZERO_; WEIGHT TOTVISWT;
Stata	svyset HC_ID [pweight= TOTVISWT], strata(STRATUM_S) fpc(POPHC)
R	#Using the package name 'survey': {variable name} <- svydesign(ids = ~ HC_ID, strata = ~ STRATUM_S, weights = ~ TOTVISWT, fpc = ~ POPHC, data = {input data frame})

NOTE: replace curly brackets {} with the information named in the parentheses.

NOTE: _ZERO_ in the TOTCNT statement of the SUDAAN example indicates that there is no visit sampling, and therefore no sampling variance within the HC_ID.

When using SUDAAN software, sort the file in the order specified by the NEST statement. For example, the file records must be sorted first by STRATUM_S and HC_ID in the above example. Below are definitions of the variables included in the above examples:

- **STRATUM_S**: Scrambled stratum value, corresponding to the original stratum from which the health center was sampled.
- **HC_ID**: the sample identifier for the health center. Use the HC_ID from the weight table when running weighted estimates, as it must be in numeric format for SAS-callable SUDAAN statements.
- **POPHC**: estimated total number of in scope HCs in the sampling stratum (STRATUM_S) from which the health center was selected.
- **TOTVISWT**: weighted value of the visit.

Analytic Requirements

As mentioned above, some health centers did not provide certain data elements for any of their visits in the 2023 data year. In certain situations, some health centers needed to produce custom extracts of their records to conform with the format needed for processing as specified in the IG. Therefore, not all data elements were required of health centers providing custom extracts. In other situations, even for health centers providing data via the IG, certain variables were incomplete for all visits at specific health centers.

Regardless of the reason for missingness, data users must identify health centers that have complete missingness for specific analytic variable(s) of interest and exclude those health centers' visits from analysis. Additionally, if certain health centers' visits must be excluded, users must normalize the weight variable (TOTVISWT) so that the sum of weights of visits in the analysis is equal to the sum of weights of all visits in the 2023 NAMCS HC Component RDC database.

Steps for complete case analysis

1. Identify health centers to be included in your analysis:
    a. Identify variable(s) required for your analysis.
    b. Identify health centers that are missing values at ALL visits for at least one variable of interest from Step 1a.
    c. Exclude all visits from health centers identified with complete missingness for at least one variable of interest, as identified in Step 1b above.

    NOTE: this process does not eliminate all missingness, rather it eliminates complete missingness of a specific variable for a specific health center. Health centers that are included may still have some visits with missing information for the variables of interest, but this process removes visits at health centers that did not provide any information for variables of interest.

2. Normalize weights with the subset of health centers' visits included in your analysis:
    a. Calculate the sum of weights for all visits in the database. In 2023, the sum of weights (TOTVISWT) is 124,293,279.
    b. Calculate the sum of weights for visits at health centers to be included in your analysis.
    c. Calculate the normalization factor [X] by dividing the sum of weights for all visits in the survey by the sum of weights for visits in your analysis, and the value of X from this calculation is the factor you will use to normalize your weights.
        i. X = [sum of all visit weights] / [sum of visit weights in your analysis]
            1. NOTE: X will always be greater than 1.
    d. Create a new weight variable in your analysis by multiplying the original weight variable by your normalization factor (X).
        i. NEW_WT = TOTVISWT * X
    e. Use NEW_WT for your analysis in place of TOTVISWT, according to the coding examples provided above. Apply this new weight variable to the subset of visits in your analysis.

NOTE: If you add or subtract variables from your analysis, or you develop a new research question and analysis, you must conduct these steps again to ensure that you: 1) capture visits from health centers' providing data on your variables of interest, and 2) normalize those visits' weights accordingly.

Examples of Steps for a Complete Case Analysis

The examples below will showcase the differences in estimates when normalizing the 2023 NAMCS HC Component RDC database for visits with a mental health disorder overall and by race. The examples will provide context on normalizing weights when assessing complete missingness for one variable (Condition) and complete missingness for multiple variables (Condition and Race).

Before following the steps for a complete case analysis, it is helpful to assess the unweighted and weighted number of visits for all 95 health centers included in the 2023 NAMCS HC Component RDC database, as shown in Table 1. There are 9,012,885 visits in the database representing a weighted value of 124,293,279 health center visits.

Table 1. Weighted and unweighted number of visits in the 2023 NAMCS HC Component RDC database

	Visits at all health centers (N=95)
Unweighted	9,012,885
Weighted	124,293,279

Source: 2023 NAMCS Health Center Component

Example 1: Normalization analysis using the condition variable

In this example, assume that the user wants to assess the count and percent of visits with a mental health disorder using the 2023 NAMCS HC Component RDC database.

- NOTE: For the purposes of this example, a mental health disorder was classified as any ICD-10-CM code in the Mental, Behavioral and Neurodevelopmental disorders chapter (F01-F99), which could be in any condition record in the database for a given visit.

First, in the condition table, the user must identify the number of health centers that have complete missingness for the variable of interest. In the condition table, CONDITION_CODE_R contains ICD-10-CM diagnosis information. In 2023, four out of the 95 health centers have complete missingness in the CONDITION_CODE_R variable. Therefore, these four health centers should be excluded from the analysis, meaning the analysis will only include visits from the 91 health centers that have any condition codes in CONDITION_CODE_R. Next, the normalization factor X should be calculated by dividing the sum of all visit weights (124,293,279) by the sum of visit weights from the 91 health centers included in the analysis (120,171,832). The normalization factor is 124,293,279/120,171,832 or approximately 1.03. As described above, the normalization factor is used to create a new weight variable, which in this example is calculated as NEW_WT = TOTVISWT * 1.03. After calculating the normalization factor and creating a new weight variable, the data user should conduct their analysis using the new visit weight variable and the subset of visits at the 91 health centers. The total sum of weights in the analytic subset of visits (normalized weighted denominator) should be equal to the total sum of weights for all visits at all health centers as seen in Table 1.

At the 91 health centers identified for inclusion in this example, we identified visits with a mental health ICD-10-CM diagnosis in any condition record. We then produced unweighted and weighted estimates (using the normalized NEW_WT variable) of visits with a mental health diagnosis at health centers in 2023. These estimates are detailed in Table 2 for users to replicate. Please note, normalization only impacts the weighted numerator and weighted denominator estimates; the unweighted counts and the weighted percentages will not change due to weight normalization.

Table 2. Visits with a Mental Health Diagnosis at Health Centers that provided Condition Codes

	Visits at health centers with any CONDITION_CODE_R values (N=91)	
	Non-Normalized	Normalized
Unweighted numerator	1,607,378	1,607,378
Unweighted denominator	8,800,815	8,800,815
Weighted numerator	21,837,692	22,586,643
Weighted denominator	120,171,832	124,293,279
Weighted Percent (Standard Error)	18.17 (2.23)	18.17 (2.23)

Source: 2023 NAMCS Health Center Component

Example 2: Normalization analysis using condition and race variables

In this example, assume the user wants to assess visits with a mental health disorder using the 2023 NAMCS HC Component RDC database, but further stratified by race.

Because different health centers may have complete missingness for condition and race information, the user must reconduct the steps for complete case analysis outlined above. In 2023, 22 health centers have complete missingness in the CONDITION_CODE_R and PATIENT_RACE_R variables; 4 health centers are missing CONDITION_CODE_R at all visits (see example 1 above), and an additional 18 health centers are missing PATIENT_RACE_R at all visits. Therefore 73 health centers make up the subset of data to analyze mental health conditions by race. In this example, the normalization factor X should be calculated by dividing the sum of all visit weights (124,293,279) by the sum of visit weights from the 73 health centers included in this example (104,674,889). The normalization factor is 124,293,279/104,674,889 or approximately 1.19. The normalization factor is used to create a new weight variable, which for this example is calculated as NEW_WT = TOTVISWT * 1.19. After calculating the normalization factor and creating a new weight variable, the data user should apply the new visit weight variable to the subset of visits at the 73 health centers to be included. The total sum of weights in the data subset (Normalized weighted denominator) should be equal to the total sum of weights for all visits at all health centers shown in Table 1.

As shown in Table 3, the examples provided above assess visits with a mental health diagnosis using two different subsets of health centers' visits. Depending on the variables of interest to the user, the weighted estimates of visits for a given condition may differ. When analyzing visits at the 91 health centers that provided condition codes, the normalized weighted number of visits with a mental health diagnosis represented 22,586,643 visits in the 2023 NAMCS HC Component RDC database. When analyzing visits at the 73 health centers that provided condition and race information, the normalized weighted number of visits with a mental health diagnosis reduced to 21,430,936 visits, despite both analyses yielding the same normalized weighted denominator.

Table 3. Visits with a mental health diagnosis for two different subsets of Health Centers

	Visits at health centers with any CONDITION_CODE_R values (N=91)		Visits at health centers with any CONDITION_CODE_R and PATIENT_RACE_R values (N=73)	
	Non-Normalized	Normalized	Non- Normalized	Normalized
Unweighted numerator	1,607,378	1,607,378	1,341,834	1,341,834
Unweighted denominator	8,800,815	8,800,815	7,720,248	7,720,248
Weighted numerator	21,837,692	22,586,643	18,048,287	21,430,936
Weighted denominator	120,171,832	124,293,279	104,674,889	124,293,279
Weighted Percent (Standard Error)	18.17 (2.23)	18.17 (2.23)	17.24 (2.23)	17.24 (2.23)

Source: 2023 NAMCS Health Center Component

In short, normalizing weights may produce different estimates when analyzing the 2023 NAMCS HC Component RDC database, due to the number of health centers that are included in the analysis. Data users should consider the full scope of their research question to make decisions on the subset of health centers to include and how normalizing visit weights may impact the estimates provided.

Tips

Data users may reference Table 4 and Table 5 to ensure that the correct number of health centers are included in their analysis when normalizing weights in a complete case analysis. For example, if a single variable is used such as Age or Sex from the Patient table, then all 95 health centers can be included in the analysis. If a combination of variables are to be used in the analysis such as Condition, Age, and Sex then 91 health centers should be included in the analysis. Data users must identify the 4 health centers that have complete missingness for at least one of the measures of interest, subset the database for analysis, and conduct the steps described above to normalize the visit weights for a complete case analysis.

Table 4. Variables that contain health centers with complete missingness in the 2023 NAMCS HC Component RDC database

2023 NAMCS HC Component RDC database table	Variable	Number of health centers included
Patient	PATIENT_ETHNICITY_R	77
Patient	MARITAL_STATUS_R	81
Patient	PATIENT_RACE_R	77
Condition	CONDITION_CODE_R	91
Condition	DIAGNOSIS_TYPE	91
Condition	CONDITION_STATUS	91
Procedure	PROCEDURE_CODESYS_NAME_R	95
Vital	VITAL_TYPE_R	95
Vital	VITAL_UNIT_R	93
Medication	MEDICATION_CODE_R	79
Lab	LAB_TESTCODE_R	77

Table 5. Number of Health Centers to include among select combinations of 2023 NAMCS HC Component RDC database tables

2023 NAMCS HC Component RDC database tables	Variables	Number of health centers included
Condition + Patient	CONDITION_CODE_R, PATIENT_AGE_D	91
Condition + Patient	CONDITION_CODE_R, PATIENT_SEX_D	91
Condition + Patient	CONDITION_CODE_R, PATIENT_AGE_D, PATIENT_SEX_D	91
Condition + Patient	CONDITION_CODE_R, PATIENT_ETHNICITY_R	73
Condition + Patient	CONDITION_CODE_R, PATIENT_RACE_R, PATIENT_ETHNICITY_R	73
Condition + Patient	CONDITION_CODE_R, PATIENT_RACE_R	73
Condition + Patient	CONDITION_CODE_R, PATIENT_AGE_D, PATIENT_SEX_D, PATIENT_RACE_R, PATIENT_ETHNICITY_R	73

NOTE: This list is not exhaustive, as there may be other combinations of variables and tables that data users may utilize in the 2023 NAMCS HC Component RDC database.