

National Center for Health Statistics Research Data Center National Ambulatory Medical Care Survey Health Center Component 2021 Restricted Use File Data Dictionary

This document contains the data dictionary for the 2021 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.

Participating HCs were asked to submit electronic health record (EHR) data for all encounters in calendar year 2021 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 2021, 29 of 111 FQHCs submitted data on 3,543,927 visits from January 1, 2021 to December 31, 2021, for an unweighted response rate of 26.1% and a weighted response rate of 26.8%. 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 1 in the AAPOR Response Rate Calculator 4.1.

The 2021 NAMCS HC Component data are weighted and can be used to produce nationally representative estimates of visits at FQHCs. However, due to the low response rate in 2021 and the resulting small number of responding FQHCs (n=29), weighted analyses

produced estimates with high variance and frequent unreliability according to 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. As a result, NCHS did not release a public use file for the 2021 NAMCS HC, but weight and design variables are being made available to users in the RDC. Users should pay close attention to variance estimates and NCHS presentation standards 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
Weight	Visit weights and design variables to produce national estimates	Visit

The table below shows the join relationships between the tables.

- $1 \rightarrow 1$ indicates one-to-one visit links.
- $1 \rightarrow \infty$ indicates one-to-many visit to observation links.
- $\infty \rightarrow \infty$ indicates many-to-many observation to observation links.

Table Name	Visit	Patient	Condition	Procedure	Weight
Visit		$1 \rightarrow 1$	$1 \rightarrow \infty$	$1 \rightarrow \infty$	$1 \rightarrow 1$
Patient	$1 \rightarrow 1$		$1 \rightarrow \infty$	$1 \rightarrow \infty$	$1 \rightarrow 1$
Condition	$\infty \rightarrow 1$	$\infty \rightarrow 1$		$\infty \rightarrow \infty$	$\infty \rightarrow 1$
Procedure	$\infty \rightarrow 1$	$\infty \rightarrow 1$	$\infty \rightarrow \infty$		$\infty \rightarrow 1$
Weight	$1 \rightarrow 1$	$1 \rightarrow 1$	$1 \rightarrow \infty$	$1 \rightarrow \infty$	

Contents

Visit Table.....	4
Patient Table.....	5
Condition Table.....	7
Procedure Table.....	9
Weight Table.....	10
Coding Examples for Weighted Estimates.....	11
Analytic Requirements.....	12
Steps for complete case analysis.....	12
Examples of Steps for a Complete Case Analysis.....	13
Example 1: Normalization analysis using the condition variable.....	13
Example 2: Normalization analysis using condition and race variables.....	14
Tips.....	15

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 HCs needed to produce custom extracts of their records to conform with the format needed for processing via the IG. These HCs' 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 "01JAN2021 : 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 an 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. The variable PATIENT_AGE_UNIT_D must be used with this variable. A small number of visits are missing patient age, and a small number of ages appear to be incorrect. 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		<p>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 HC visits.</p> <p>In this variable, periods have been removed from the 4th position or the value; for instance, "F17.200" (Nicotine dependence, unspecified, uncomplicated) is represented as "F17200".</p> <p>CONDITION_CODE_R is missing from 14% of condition records, and about 60% 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 for your analysis.</p>
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	<p>“Primary”</p> <p>“Diagnosis”</p> <p>“” = Missing</p>	<p>“Primary” diagnosis is identified by the presence of LOINC code 52534-5, as specified in the IG.</p> <p>Diagnosis type is missing in 49% of condition records, so DIAGNOSIS_TYPE may not be particularly useful.</p>
CONDITION_STATUS	Indicator for whether a condition code is active or not	Character	<p>“Active”</p> <p>“Inactive”</p> <p>“” = Missing</p>	<p>“Active” versus “Inactive” conditions are identified by SNOMED CT codes of 55561003 and 73425007, respectively, as specified in the IG.</p> <p>CONDITION_STATUS is missing in 5% of condition records.</p>

Procedure Table

RDC Variable Name	Description	Format	Value	Notes
VISIT_ID	NCHS-assigned visit identifier	Numeric		Key variable
PROCEDURE_CODE	Procedure code	Character		<p>Only 16% of visits have at least one procedure code record. However, all procedure codes for a given visit are included as originally extracted from the EHR record in this PROCEDURE_CODE variable. Multiple code systems are represented as seen in the PROCEDURE_CODESYS_NAME variable below.</p> <p>Note, 6% of records in the procedure table are missing a value for PROCEDURE_CODE, and 16% of records have values that were extracted as "procedure[i]" and "OTH" for PROCEDURE_CODE.</p>
PROCEDURE_CODESYS_NAME	Procedure code system name	Character	"CPT" "CPT4" "SNOMED CT" "" = Missing	

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 HCs in the sampling stratum from which the HC was selected	Numeric		
STRATUM_S	Scrambled sampling stratum from which the HC 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 HC was sampled.
- **HC_ID**: the sample identifier for the HC. 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 HC 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 2021 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 2021 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 2021, the sum of weights (TOTVISWT) is 120,272,406.
 - 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. $X = [\text{sum of all visit weights}] / [\text{sum of visit weights in your analysis}]$
 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. $\text{NEW_WT} = \text{TOTVISWT} * X$
 - 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 2021 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 29 health centers included in the 2021 NAMCS HC Component RDC database, as shown in Table 1. There are 3,543,927 visits in the database representing a weighted value of 120,272,406 health center visits.

Table 1. Weighted and unweighted number of visits in the 2021 NAMCS HC Component RDC database

	Visits at all health centers (N=29)
Unweighted	3,543,927
Weighted	120,272,406

Source: 2021 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 2021 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 2021, one out of the 29 health centers have complete missingness in the CONDITION_CODE_R variable. Therefore, these six health centers should be excluded from the analysis, meaning the analysis will only include visits from the 28 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 (120,272,406) by the sum of visit weights from the 28 health centers included in the analysis (118,164,113). The normalization factor is $120,272,406/118,164,113$ or approximately 1.02. As described above, the normalization factor is used to create a new weight variable, which in this example is calculated as $NEW_WT = TOTVISWT * 1.02$. 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 28

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 28 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 2021. 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=28)	
	Non-Normalized	Normalized
Unweighted numerator	449,622	449,622
Unweighted denominator	3,471,232	3,471,232
Weighted numerator	17,437,078	17,748,192
Weighted denominator	118,164,113	120,272,406
Weighted Percent (Standard Error)	14.76 (2.30)	14.76 (2.30)

Source: 2021 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 2021 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 2021, two health centers have complete missingness in the CONDITION_CODE_R and PATIENT_RACE_R variables; one health center is missing CONDITION_CODE_R at all visits (see example 1 above), and one additional health center is missing PATIENT_RACE_R at all visits. Therefore 27 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 (120,272,406) by the sum of visit weights from the 27 health centers included in this example (113,346,162). The normalization factor is $120,272,406/113,346,162$ or approximately 1.06. The normalization factor is used to create a new weight variable, which for this example is calculated as $NEW_WT = TOTVISWT * 1.06$. 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 27

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 28 health centers that provided condition codes, the normalized weighted number of visits with a mental health diagnosis represented 17,748,192 visits in the 2021 NAMCS HC Component RDC database. When analyzing visits at the 27 health centers that provided condition and race information, the normalized weighted number of visits with a mental health diagnosis reduced to 15,925,856 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=28)		Visits at health centers with any CONDITION_CODE_R and PATIENT_RACE_R values (N=27)	
	Non-Normalized	Normalized	Non- Normalized	Normalized
Unweighted numerator	449,622	449,622	414,357	414,357
Unweighted denominator	3,471,232	3,471,232	3,401,265	3,401,265
Weighted numerator	17,437,078	17,748,192	15,008,719	15,925,856
Weighted denominator	118,164,113	120,272,406	113,346,162	120,272,406
Weighted Percent (Standard Error)	14.76 (2.30)	14.76 (2.30)	13.24 (2.08)	13.24 (2.08)

Source: 2021 NAMCS Health Center Component

In short, normalizing weights may produce different estimates when analyzing the 2021 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 29 health centers can be included in the analysis. If a combination of variables are to be used in the analysis such as Condition, Age, Sex, Race, Ethnicity, and Marital Status then 25 health centers should be included in the analysis. Data users must identify the 4 health centers that have complete missingness for at least on 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 2021 NAMCS HC Component RDC database

2021 NAMCS HC Component RDC database table	Variable	Number of health centers included
Patient	MARITAL_STATUS_R	27
Patient	PATIENT_RACE_R	28
Condition	CONDITION_CODE_R	28
Condition	DIAGNOSIS_TYPE	25
Condition	CONDITION_STATUS	25
Procedure	PROCEDURE_CODESYS_NAME	28

Table 5. Number of Health Centers to include among select combinations of 2021 NAMCS HC Component RDC database tables

2021 NAMCS HC Component RDC database tables	Variables	Number of health centers included
Condition + Patient	CONDITION_CODE_R, PATIENT_AGE_D	28
Condition + Patient	CONDITION_CODE_R, PATIENT_SEX_D	28
Condition + Patient	CONDITION_CODE_R, PATIENT_AGE_D, PATIENT_SEX_D	28
Condition + Patient	CONDITION_CODE_R, PATIENT_ETHNICITY_R	28
Condition + Patient	CONDITION_CODE_R, PATIENT_RACE_R, PATIENT_ETHNICITY_R	27
Condition + Patient	CONDITION_CODE_R, PATIENT_RACE_R	27
Condition + Patient	CONDITION_CODE_R, PATIENT_AGE_D, PATIENT_SEX_D, PATIENT_RACE_R, PATIENT_ETHNICITY_R, MARITAL_STATUS_R	25

NOTE: This list is not exhaustive, as there may be other combinations of variables and tables that data users may utilize in the 2021 NAMCS HC Component RDC database.