# 2022 NATIONAL AMBULATORY MEDICAL CARE SURVEY HEALTH CENTER (NAMCS HC) COMPONENT TECHNICAL DOCUMENTATION

For Public Use Data File

**Division of Health Care Statistics**
**National Center for Health Statistics**
**May 2024**

## Overview Summary

This document provides detailed information and guidance for users of the 2022 National Ambulatory Medical Care Survey Health Center (NAMCS HC) Component public use data file. As a principal source of information on health care utilization in the United States, the NAMCS HC Component collects visit data from a nationally representative sample of U.S. federally qualified health centers (FQHCs) and FQHC look-alikes through electronic health record (EHR) data submission. The 2022 NAMCS HC Component is conducted by the National Center for Health Statistics (NCHS) and is a member of the National Health Care Surveys – a family of surveys which measure health care utilization across a variety of health care providers and settings.

Section 1 of this document includes information on the scope of the survey, the data sources, and the confidentiality protections related to the data. Section 2 contains details on the sampling process, data collection procedures, and weighting methodology used to produce national estimates on health care utilization. Section 3 provides information on the number of sampled health centers that were eligible to participate in the NAMCS HC Component and submitted data in 2022. Section 4 details the contents of the 2022 NAMCS HC Component public use data file and the edits used in the creation of the file. Section 5 contains an explanation of the procedures used to accurately produce variance estimates. NCHS presentation standards for proportions, counts, and rates, and their relation to NAMCS HC Component data, are discussed in Section 6, and the data analysis guidelines are provided in Section 7. Section 8 provides information on item missingness, and Section 9 provides a comparison of frequencies between the NAMCS HC Component public use and restricted use data files. Section 10 provides a list of preferred reporting items for complex sample survey analysis. Section 11 provides further information on the availability of NAMCS HC Component restricted use data files available in NCHS and Federal Research Data Centers. Appendix A provides unweighted frequencies for selected variables included on the public use data file.

## Suggested Citation

Technical Documentation: National Center for Health Statistics. Division of Health Care Statistics. *2022 National Ambulatory Medical Care Survey Health Center (NAMCS HC) Component Public Use Data File Documentation,* May 2024. Hyattsville, Maryland.

Data File: National Center for Health Statistics. Division of Health Care Statistics. 2022 National Ambulatory Medical Care Survey Health Center (NAMCS HC) Component public use data file. 2024. Hyattsville, Maryland.

## Contact Information

Data users can find the latest information about the NAMCS HC Component on our website, at: https://www.cdc.gov/nchs/ahcd/namcs_index.htm. If data users have queries about the public use data file, they may send their question through email to ambcare@cdc.gov, or call us at 301-458-4600. A response to data user inquiries is generally provided in 1-2 business days.

The National Center for Health Statistics has an ambulatory health care data listserv, where updates and information about the most recent ambulatory care data (including the NAMCS HC Component) are sent out. Details on how to subscribe to the NCHS Listserv for ambulatory health care data can be found at: https://www.cdc.gov/nchs/ahcd/ahcd_listserv.htm.

# Contents

# Section 1 About the National Ambulatory Medical Care Survey Health Center Component

## Section 1.1 Background

The National Ambulatory Medical Care Survey Health Center (NAMCS HC) Component is an annual survey that provides data on health care utilization at health centers in the United States. As a part of NAMCS, the National Center for Health Statistics (NCHS) began collecting data from health centers in 2006. A separate sample of health centers was drawn in 2012 for NAMCS. In 2021, NCHS redesigned the NAMCS HC Component to collect visit data from electronic health records (EHRs) from participating health centers for the entire calendar year.

The NAMCS HC Component collects data on health center visits including information on diagnoses and patient demographics. The survey aims to provide health trends and outcomes of the U.S. population's utilization of health centers in the following ways:

- Provide nationally representative, accurate, and reliable health care data for health centers in the United States.
- Answer key questions of interest to health care professionals, researchers, and policy makers about health care quality, use of resources, and disparities of services to population subgroups.
- Monitor national trends in health care topics for which health centers play an important role, such as mental health and substance use-related care, maternal and child health, and HIV-related care.
- Contribute to a stronger public health foundation that helps address current and future public health threats.

In 2022, the entire sample included 324 federally qualified health centers (FQHC) and FQHC look-alikes in the 50 U.S. states and the District of Columbia that used an EHR system. Out of the entire sample, 104 health centers were included in the primary sample and 220 health centers made up the reserve sample. Ultimately, 255 health centers were contacted and 64 health centers agreed to participate and provided visit data from their EHRs. Out of the 64 responding health centers, 26 continued participation from 2021 and 38 health centers were newly recruited in 2022. For more detailed information regarding the sample frame, see Section 2.2.

Overall, 5,640,370 health centers visits were collected from the 64 responding health centers. Of these, 282,017 health center visits were selected to create the 2022 NAMCS HC Component public use data file.

## Section 1.2 Data Sources

The NAMCS HC Component receives data from EHR systems. Participating health centers submit EHR data, which contain an unlimited number of medical diagnosis and procedure codes, laboratory and medication data, and unstructured clinical notes. However, the public use data file will only include diagnosis variables and demographic information. The NAMCS HC Component accepts EHR data in the format of HL7 CDA® R2 Implementation Guide: National Health Care Surveys Release 1, DSTU Release 1.2 – US Realm (http://www.hl7.org/implement/standards/product_brief.cfm?product_id=385). However, some EHR vendors are not able to format their data in the HL7 CDA format as specified in the National Health Care Surveys Implementation Guide. Alternatively, these centers were able to submit their EHR data as custom extracts, which contained many (but not all) data elements extracted via the above implementation guide.

## Section 1.3 Data Confidentiality

NCHS and its agents take the security and confidentiality of NAMCS HC Component public use data file very seriously. Strict laws have been implemented to establish minimum Federal standards for safeguarding the privacy of individually identifiable health information. Assurance of confidentiality is provided to all health centers according to Section 308(d) of the Public Health Services Act [42 United States Code 242m (d)]. Strict procedures according to Section 3572 of the Confidential Information Protection and Statistical Efficiency Act (44 U.S.C. 3561-3583) are utilized to prevent disclosure of personal identifiable information in NAMCS HC Component data. All information which could identify a participating health center is confidential and seen only by persons associated with NAMCS HC Component, and is not disclosed or released to others for any other purpose. Prior to the release of public use data file, NCHS conducts extensive disclosure risk analysis to minimize the chance of inadvertent disclosure. As a result, selected characteristics and/or data elements may have been omitted or masked on the public use data file to minimize the potential risk of disclosure. Masking was performed in such a way to cause minimal impact on the data. See Section 4: Data Processing for more information on which data elements in the public use data file were impacted.

The protocol for NAMCS HC Component has been approved by the NCHS Research Ethics Review Board since the survey's establishment (2006).

# Section 2 Methodology

## Section 2.1 Brief Overview

The 2022 NAMCS HC Component used a national probability sample of health centers to collect data on visits to develop the public use data file. The 2022 NAMCS HC Component public use data file sample was designed to allow for nationally representative estimates of visits at health centers in the United States.

## Section 2.2 Health Center Frame and Sample Design

The 2022 NAMCS HC Component identified a targeted universe of FQHCs and FQHC look-alikes in the 50 U.S. states and the District of Columbia that provide direct ambulatory care and use an EHR system at one or more delivery sites. Health centers that were fully or partially funded by the Health Resources and Services Administration (HRSA) were considered for inclusion. Health centers were deemed ineligible if they:

- Did not have an EHR system
- Did not provide healthcare services to the general U.S. population, or only provided care to special institutionalized populations such in prisons, nursing homes, homeless shelters, etc.
- Only provided dental services
- Were located on a military installation or outside of the 50 U.S. states and the District of Columbia

To create the sampling frame and draw the sample, NCHS worked with the HRSA to use a nationally representative database that contains a list of all health centers in the United States. The database contained 1,482 health centers for the 2022 NAMCS HC Component. To create the sampling frame from this database, ineligible health centers were removed. This included 64 health centers that did not meet the inclusion criteria described above and 149 health centers that were included in the 2021 sample. This process yielded a sampling frame of 1,269 eligible health centers.

In 2021, a stratified random sample of 50 FQHCs and FQHC look-alikes was drawn as the primary sample, along with a reserve sample of 100 health centers. The 2022 NAMCS HC Component sample was expanded to initially add 60 respondent health centers to the 50 respondent health centers from the

2021 sample, resulting in 110 FQHCs and FQHC look-alikes making up the 2022 NAMCS HC Component sample. However, 54 health centers were ultimately fielded due to budget constraints. Due to this, six randomly selected health centers were removed from the sample in four strata. In 2022, an additional 120 additional health centers were selected for the reserve sample (Williams et al., 2023).

Ultimately 255 health centers were contacted to participate in the 2022 NAMCS HC Component, which includes 64 respondents and 191 eligible non-respondents. The 64 participating health centers include 26 health centers from the 2021 sample and 38 health centers from the 2022 sample. Weighting was conducted to produce health center-level and visit-level estimates. Data were collected for 100% of visits from the sampled health centers via EHR submission.

## Section 2.3 NAMCS HC Component Public Use Data File Sample Design

While the NAMCS HC Component restricted use data file includes every health center ($HC_j$) visit record submitted to NAMCS HC Component for the survey year, the 2022 NAMCS HC Component public use data file consists of a 5% sample of NAMCS HC Component visit data. This 5% sample of NAMCS HC Component records was selected for the public use data file instead of the full listing of records to decrease disclosure risk and increase efficiency for data users when conducting statistical analyses.

In 2022, the NAMCS HC Component collected 5,640,370 visit records. Stratified systematic sampling was used to select the public use data file sample of health center visits. A targeted number of records was determined by taking 5% of the total health center visit records (n=282,017). The sampling interval was the inverse of the percent of submitted EHRs targeted for inclusion in the subsample. The sampling interval used to select the public use data file records in the 2022 NAMCS HC Component was 1/0.05, or 20. Within each estimation stratum, participating health centers were randomly ordered. Within each health center, visits were then sorted by the following variables:

Visit Week → Day of Week

Once sorted, visits were serially numbered in each estimation stratum. Next from the ordered array of $HC_j$ records, visits were selected for the public use data file sample if the assigned "array sequential" numbers were the nearest integer greater than or equal to:

$$R_\Gamma + Int(EHR)_\Gamma \times k$$

Where:

$\boldsymbol{R}_\Gamma$ = random number between 0 and $Int(EHR)_\Gamma$

$\boldsymbol{k}$ = 0, 1, 2, 3….

$\boldsymbol{Int(EHR)_{\Gamma}}$ = sampling interval

## Section 2.4 Data Collection Procedures

In 2022, health centers submitted EHR data via two sources, either directly from the health centers' EHR system or as a custom extract, as mentioned above in Section 1.2. Once data were collected, several steps were required for data processing. Specifications for checking, configuring, and transmitting the data files were developed by NCHS. Once NCHS received the data files they were processed to harmonize data from the two data sources. All records from participating health centers' EHRs were brought into the restricted database, and those records were then collapsed so that a given patient could only have one record (called a visit in the PUF) per day at a given health center.

## Section 2.5 Weighting

Weighting was conducted to produce health center-level and visit-level estimates, and to account for sampling probabilities and nonresponse. Only visit-level weights are included in the public use data file, and users are only able to produce visit-level estimates with this file.

Health center-level data were collected via self-completed forms from participating health centers. All 2022 health center visits were collected from the sampled health centers via electronic files of their EHR system. Participating health centers submitted data for all visits that occurred during the 2022 calendar year. While the 2022 NAMCS HC Component restricted use data file includes all (100%) of the visit records sent, the public use data file includes a 5% sample of those records, as described in Section 2.3.

All health center visit data collected for 2022 were used to develop weights. To produce visit-level weights, health center-level weights were first developed and smoothed. The visit-level weights were then developed for the restricted use file that includes all visits from participating health centers. These visit weights were formulated as the final health center weight multiplied first by the health center's actual annual number of visits made for medical care followed by a partial non-response adjustment factor. Visit weights for all visits were then smoothed before they were finalized. Because the public use data file only contains a 5% sample of all visits submitted in the 2022 NAMCS HC Component, visit weights for visits included on the public use data file were adjusted accordingly. This ensures that weighted estimates from the restricted use file and the public use data file sum to approximately the same total number of weighted visits at health centers in the survey year.

Variance estimation procedures for weighted estimates are described further in Section 5 with coding examples in Section 7, and comparisons of weighted estimates between the restricted and public use data files in Section 9.

# Section 3 Sample Size, Eligibility, and Response Rate

All 255 health centers that were contacted for participation were eligible to participate in the survey. Ultimately, 64 health centers participated in the 2022 NAMCS HC Component yielding a response rate of 25.1%. As described in Section 2.2, 54 health centers in 2022 were added to the 50 health centers selected in 2021, totaling to 104 health centers in the 2022 NAMCS HC Component sample. With this target of recruiting and securing 104 health centers to participate in the 2022 NAMCS HC Component, 64 ultimately participated (or 61.5% of this targeted goal) ultimately agreed. A health center was considered a full respondent if they provided data for at least six months of the survey year. Of the 64 participating health centers that were included in the 2022 NAMCS HC restricted use data file, all provided at least six months of data. Therefore, all health centers were selected to create the public use data file. From the 64 health centers, 5% of all records were selected for the public use data file. Overall, 282,017 health center visits were selected. Table 3.1 presents the number of health centers, visits, and response rates for the 2022 NAMCS HC Component.

**Table 3.1 Number of health centers, visits, and unweighted response rates, NAMCS HC Component, 2022**

| | TOTAL | | |
|---|---|---|---|
| | Health Centers | Visits | Unweighted Response Rate* |
| **Restricted Use Data File** | 64 | 5,640,370 | 25.1 |
| **Public Use Data File** | 64 | 282,017 | N/A |

**Note:** N/A is not applicable.

*Response Rate was calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 formula. The percentage is a calculation of the eligible respondents and partial respondents (N=64) divided by the eligible respondents, partial respondents, eligible non-responding and not contacted respondents (N=255).

# Section 4 Data Processing

The data included in the public use data file underwent additional processing to prepare them for release. Suppression rules such as masking were applied for some records to protect patient confidentiality. Other items were either top-coded or bottom-coded in accordance with NCHS confidentiality requirements; this is noted for specific data items outlined in this section. Imputation was not conducted for data elements with missing values prior to creation of the 2022 NAMCS HC Component public use data file.

## Section 4.1 Diagnosis Data

In the 2022 NAMCS HC Component, diagnosis data from participating health centers were submitted in three different diagnosis coding systems including: *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*; *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)*; and *SNOMED Clinical Terms (SNOMED CT)*. In the creation of a harmonized and integrated database, the ICD-9-CM and SNOMED CT diagnosis codes were translated to ICD-10-CM, where applicable. Translation from ICD-9-CM and SNOMED CT to ICD-10-CM was the only modification to the diagnosis codes. On the public use data file, medical diagnosis codes were limited to ICD-10-CM diagnosis codes.

An ICD-10-CM code can have a maximum of 7 characters and is organized by chapters from A to Z. For the 2022 NAMCS HC Component public use data file, ICD-10-CM codes have been truncated to four characters to minimize disclosure risks. While the codes have been truncated, the diagnosis codes are never updated or revised to a different code that would result in a change to the original diagnosis for a visit. To maintain integrity of the data, any codes that appear to be invalid are kept as is.

Duplicate 4-character ICD-10-CM codes were removed for each unique visit on the public use data file. Although visits collected from health center EHR systems could have had an unlimited number of diagnosis records, diagnosis codes were limited to 30 unique codes per visit (variables DX1 through DX30) in the public use data file, which captured 96.6% of diagnoses recorded at visits included on the public use data file. Rarity of diagnoses was assessed and those deemed rare were truncated to two characters.

At least one diagnosis code is listed in 62.0% of all visits. Six health centers did not provide any condition codes that could be translated to ICD-10-CM, therefore do not have any visits that include at least one

condition code in DX1-DX30. Of the 58 health centers that provide any codes that translated to ICD-10-CM, 74.4% of their visit have at least one diagnosis code in the public use data file.

## Section 4.2 Patient Age

Patient age is present for all visits in the 2022 NAMCS HC Component public use data file. Visits were top coded to the 99.5$^{th}$ percentile of age, thus visits by patients ages 88 and older were top coded to 88 years.

## Section 4.3 Patient Sex

Patient sex is missing in 0.1% of records on the 2022 NAMSC HC Component public use data file.

## Section 4.4 Patient Race and Hispanic Ethnicity

Patient race is missing from 24.5% of records on the 2022 NAMCS HC Component public use data file. Eleven health centers are missing patient race for all visit records. Excluding the 11 health centers with complete missingness, 17.7% of visits are missing patient race.

Patient ethnicity is missing from 12.9% of records on the 2022 NAMCS HC Component public use data file. Ten health centers are missing patient ethnicity for all visit records. Excluding the 10 health centers with complete missingness, 6.6% of visits are missing patient ethnicity.

## Section 4.5 Patient Marital Status

Marital status of patients is included in the public use data file but is missing from 20.9% of records overall. Ten health centers are missing marital status from all visit records. For the remaining 54 health centers, marital status is missing from 15.2% of visits.

## Section 4.6 Visit Month and Day

Exact dates are not provided on the NAMCS HC Component public use data file. Instead, only the month and day of the week of health center visits are provided.

# Section 5 Standard Errors and Variance Estimation

Standard error is primarily a measure of the sampling variability that occurs by chance because only a sample of health centers are in NAMCS HC Component, rather than the entire universe of health centers. Standard errors and other measures of sampling variability are best determined by using a statistical software package that takes into account the sample designs of surveys to produce such measures.

See Section 7 for further guidance on how to apply weights and calculate standard errors to generate national estimates.

## Section 5.1 Subpopulation Analysis – Subsetting Data

For data users who may have a subpopulation of interest, such as a particular age group or sex, a domain analysis must be performed, also known as a subgroup or subpopulation analysis.

For some variance estimation methods, the entire set of data containing the appropriate weights for a particular survey year must be used to obtain the correct variance estimates. Therefore, it is not recommended to drop observations from the dataset when subsetting data, as it may affect variance estimation. Instead, the estimation procedure must indicate which records are in the subgroup of interest. For example, when examining female patients aged 35 and over, the entire dataset of examined individuals (both male and female patients of all reported ages) must be read into the statistical software program.

The STAT and DOMAIN statements in the SAS survey procedure, SUBPOPN in SAS callable SUDAAN, or comparable statements in other programs (SUBSET in R; subpop or over in Stata) must be used to indicate the subgroup of interest (i.e., females aged 35 and over).

Depending on the specifications of a data user's statistical software of choice, an indicator variable created by the data user prior to the procedure may facilitate the identification of the subgroup in the procedure statements.

## Section 6 Presentation Standards

Data users should be aware of the reliability of survey estimates, particularly smaller estimates. NCHS has published standards for the assessment of reliability and presentation of proportions (or percentages) (https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf) and for the presentation of rates and counts (https://www.cdc.gov/nchs/data/series/sr_02/sr02-200.pdf). For presentation or publication of count estimates using data from the NAMCS HC Component, we recommend visit estimates be rounded to the nearest thousand.

These presentation standards apply to products published by NCHS. If, according to the presentation standards, an estimate is not reliable, data users should examine the confidence interval carefully before using the estimate.

# Section 7 Data Analysis Guidance

The following section provides an overview on how data users can derive visit estimates and compute variances to produce standard errors, using statistical software tools such as SAS, R, and Stata. For the NAMCS HC Component public use data file, SAS-callable SUDAAN software procedures are used for survey analysis, however, SAS/STAT software procedures beginning with SURVEY for survey analysis may also be used. R relies on the "survey" package to conduct survey data analysis whereas Stata, uses the "svy" command. SAS/SUDAAN, R and Stata users can use these procedures to conduct statistical analysis on data from the 2022 NAMCS HC Component public use data file. Additionally, this section provides guidance on normalizing visit weights to account for complete missingness for analytic variables of interest. The guidance provides data users a framework to implement normalizing weights for data analysis. Data users should always investigate if there are any variables of interest that have complete missingness at health centers in the 2022 NAMCS HC Component public use data file.

## Section 7.1 Visit weight

The visit weight is a critical component in the process of producing estimates from sample data and its use should be clearly understood by all data users. The statistics contained on the public use data file reflect only a sample of visits; a 5% sample of the NAMCS HC Component data collected from participating health centers, not a complete count of all visits that occurred in the United States. Each health center's visit record in the public use data file represents one patient visit in the sample of 282,017 visits. To obtain national estimates from the 5% sample, each record is assigned an inflation factor called the "visit weight" (variable VISWT in the public use data file).

By aggregating the "visit weights" assigned to the VISWT variable on the 282,017 health center visits for 2022, the data user can obtain the estimated total of 109,087,913 health center visits (standard error of 19,896,515 health center visits) made in the United States in 2022.

Note that estimates of health center visits produced from the 2022 NAMCS HC Component public use data file may differ somewhat from those estimates produced from the 2022 NAMCS HC Component restricted use data file. This is because of adjustments required for the public use data files, as part of the disclosure risk mitigation process. Certain variables were masked on some records for confidentiality purposes. Other variables were top and/or bottom coded in accordance with NCHS confidentiality requirements.

The table in Section 9 compares aggregate unweighted and weighted data for selected variables between the 2022 NAMCS HC Component public use data file and restricted use data file.

## Section 7.2 Guidance on Weight Normalization

Some health centers did not provide certain data elements for any of their visits in the 2022 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 HL7 CDA Implementation Guide. 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 (VISWT) so that the sum of weights of visits in the analysis is equal to the sum of weights of all visits in the 2022 NAMCS HC Component public use data file.

Steps for a 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. Subset 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 if only a subset of health centers' visits is included:
    a. Calculate the sum of weights for all visits in the public use data file. In 2022, the sum of weights (VISWT) is 109,087,913.
    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 (from step 2a) by the sum of weights for visits in your analysis (from Step 2b), 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 for visits in your analysis by multiplying the original weight variable by your normalization factor (X).
    i. NEW_WT = VISWT * X
e. Use NEW_WT for your analysis in place of VISWT.

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.

**Table 7.1 Variables that contain health centers with complete missingness in the 2022 NAMCS HC public use data file**

| Variable Name | Variable Description | HCID_S to exclude |
|---|---|---|
| DX1-DX30 | Diagnoses 1-30 | 22, 26, 42, 46, 60, 62 |
| ETHNICITY | Patient Hispanic ethnicity | 4, 11, 12, 18, 20, 23, 25, 30, 47, 63 |
| MARITAL | Marital status | 4, 11, 12, 18, 20, 23, 25, 30, 47, 63 |
| RACE | Patient race | 4, 11, 12, 18, 20, 23, 25, 29, 30, 47, 63 |
| RACERETH | Combined race and ethnicity variable | 4, 11, 12, 18, 20, 23, 25, 29, 30, 47, 63 |

## Section 7.2.1 Normalization Example

The example below will showcase the differences in estimates when normalizing the 2022 NAMCS HC public use data file for visits with a mental health disorder and race as opposed to not normalizing. This example will provide context on normalizing weights when assessing complete missingness for two variables on the public use data file (DX1 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 64 health centers in the public use data file, as shown in Table 7.2. There are 282,017 visits in the public use data file representing 109,087,913 health center visits.

**Table 7.2 Weighted and unweighted number of visits in the 2022 NAMCS HC Component public use data file**

| | Visits at all health centers (N=64) |
|---|---|
| Unweighted | 282,017 |
| Weighted | 109,087,913 |

In this example, assume the user wants to assess visits with a first-listed diagnosis (DX1) of a mental health disorder, stratified by race (RACE) using the 2022 NAMCS HC Component PUF. 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). Please note that in this public use data file, when DX1 is missing, all DX1-DX30 variables will be missing, so whether assessing first-listed or any-listed diagnosis, we only need to assess complete missingness for DX1.

First, the user must identify all health centers that have complete missingness in *either* the race (RACE) *or* first-listed diagnosis (DX1) variables (or both) from Table 7.1 above. In 2022, 17 health centers have complete missingness in the DX1 or RACE variables. Health centers 22, 26, 42, 46, 60, 62 are missing DX1 at all visits. Health centers 4, 11, 12, 18, 20, 23, 25, 29, 30, 47, and 63 are missing RACE at all visits. Therefore, 47 health centers make up the subset of data to analyze first-listed mental health diagnoses by race. The normalization factor X should be calculated by dividing the sum of all visit weights (109,087,913) by the sum of visit weights from the 47 health centers included in this example (74,065,859). The normalization factor is (109,087,913/74,065,859) or approximately 1.47. The normalization factor is used to create a new weight variable, which for this example is calculated as NEW_WT=VISWT*(1.47). 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 47 health centers included in this example. The total sum of weights in the analytic subset (sum of NEW_WT at HC visits to be included) should be equal to the total sum of weights for all visits at all health centers in the NAMCS HC public use data file as shown in Table 7.2.

At the 47 health centers identified for inclusion in this example, we identified visits with a first-listed mental health ICD-10-CM diagnosis and race information. We then produced unweighted and weighted estimates (using the normalized NEW_WT variable) of visits with a first-listed mental health diagnoses at health centers in 2022. These estimates are detailed in Table 7.3 for users to replicate. Please note, normalization of weights at the subset of visits to be included only impacts the weighted numerator and

weighted denominator estimates; the unweighted counts and weighted percentage will not change in the same subset of visits due to weight normalization.

**Table 7.3 Visits with a first-listed mental health diagnosis, with race and diagnosis information, in the 2022 NAMCS HC public use data file**

| | Overall | Non-Normalized subset | Normalized subset |
|---|---|---|---|
| Analysis | Overall Data File | Subset without Normalization implemented | Subset with Normalization correctly implemented |
| Number of health centers | 64 | 47 | 47 |
| Unweighted numerator | 23,642 | 21,151 | 21,151 |
| Unweighted denominator | 282,017 | 211,452 | 211,452 |
| Weight used | VISWT | VISWT | NEW_WT[1] |
| Weighted numerator | 8,958,206 | 7,888,090 | 11,617,975 |
| Weighted denominator | 109,087,913 | 74,065,859 | 109,087,913 |
| Weighted Percent (SE) | 8.21 (1.69) | 10.65 (1.58) | 10.65 (1.58) |

[1] As described in Section 7.2.1, NEW_WT= VISWT *1.47, where 1.47 is the calculated normalization factor.

In the first column of Table 7.3, the data is neither subset nor using a normalized visit weight. The weighted numerator underestimates the weighted number of visits with a first-listed mental health diagnosis and race, which also results in an underestimated weighted percent. In the second column, the data is subset to exclude health centers with complete missingness but does not use the normalized visit weight. This further underestimates the weighted number of visits with a first-listed mental health diagnosis. Additionally, because of the use of the subset of health centers and a non-normalized visit weight in the second column, the weighted denominator does not add up to the total number of visits in the public use data file. The last column displays the correct analysis using the subset of health centers and the normalized weight variable.

Using a subset of health centers and normalizing their weights produces a higher weighted numerator than using all health centers and the non-normalized weight or using the subset of health centers and the non-normalized weight. In the overall analysis in Table 7.3, visits at health centers with complete missingness for diagnosis data are automatically considered to be non-mental health visits despite not having enough information to discern whether there was a mental health diagnosis. Consequently, the overall weighted numerator is an undercount of visits with a first-listed mental health diagnosis at health centers in the United States.

In short, normalizing weights may produce different estimates when analyzing the 2022 NAMCS HC Component public use data file depending on the number of health centers that are included in the

analysis. Without excluding health centers with complete missingness and subsequently normalizing visit weights, data users will underreport counts and rates for their analysis of interest. 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 will impact the calculation of estimates. Data users should reference Table 7.1 to ensure the correct health centers are excluded in their analysis when normalizing weights in a complete case analysis.

### Section 7.3.1 Normalization Example Code

For further assistance in implementing normalization on the 2022 NAMCS HC Component public use data file, the following SAS code replicates the normalization example described in Section 7.2.1.

```
*STEP 1;
*Identify the variables of interest for your analysis;
        *Research Question: First listed diagnoses of mental health by age and race;
        *Variables needed: DX1, RACE;

*In this example you will need to subset the data where DX1 is missing or RACE
is missing according to Table 7.1;
        *DX1 is missing where HCID_S in (22, 26, 42, 46, 60, 62);
        *RACE is missing where HCID_S in (4, 11, 12, 18, 20, 23, 25, 29, 30, 47, 63);

*STEP 2;
*Calculate two sums:
        1. the sum of weights at all HCs in the original datafile and
        2. the sum of weights at HCs to be included in your analysis;
*1. Overall sum of weights;
proc sql;
        create table sum_total as
        select sum(viswt) as sum_total
        from /*[full datafile]*/;
quit;
proc print data=sum_total;
run;

*2. Subset sum of weights;
proc sql;
        create table sum_subset as
        select sum(viswt) as sum_subset
        from /*[full datafile]*/
        where HCID_S not in (4, 11, 12, 18, 20, 22, 23, 25, 26, 29, 30, 42, 46, 47, 60, 62, 63);
quit;
proc print data=sum_subset;
run;
```

```
*STEP 3;
*Create two new variables for your analysis:
        1. a normalized weight, using sum_total and sum_subset calculated in step 4 and
        2. an inclusion indicator where the record is at a PSU identified in 'all_three' from STEP 3 above;
data /*new_datafile*/;
        set /*[full datafile]*/;
                new_wt=viswt*(/*[value of sum_total]/[value of sum_subset]*/);
                if HCID_S not in (4, 11, 12, 18, 20, 22, 23, 25, 26, 29, 30, 42, 46, 47, 60, 62, 63)
                        then include=1;
                        else include=2;
                if "F01"<substr(DX1, 1, 3)<"F99"
                        then mntlhlth=1;
                        else mntlhlth=0;
run;


*STEP 4;
*Use these two variables (new_wt and include) in all procedures used to produce weighted output for
this analysis;
*Note: this step shows a SUDAAN procedure for setting up a weighted analysis, but an example of a SAS
procedure is provided below in Section 7.4;
*First, sort the data by STRATUM and HCID_S;
proc sort data=/*[new datafile]*/;
        by STRATUM_S HCID_S;
run;

*Second, set up your SUDAAN statement as follows (showing a crosstab procedure);
proc crosstab data=[new_datafile] filetype=sas design=wr atlevel1=1 atlevel2=2;
        nest STRATUM_S HCID_S / MISSUNIT;
        weight new_wt;
        subpopn include=1;
        class mntlhlth; *include analytic indicators/variables to cross;
        tables mntlhlth; *cross your class variables in the desired order;
        output nsum wsum sewgt totper setot atlev1 atlev2 / filename = /*[output dataset]*/ replace
tablecell=default;
run;
```

## Section 7.3 SAS SUDAAN Survey Procedures

The program below demonstrates how to set up your design and weight variables to produce weighted estimates using the 2022 NAMCS HC Component public use data file:

```
PROC (procedure) DATA=(input data set) FILETYPE=SAS DESIGN=WR ATLEVEL1=1 ATLEVEL2=2;
NEST STRATUM_S HCID_S / MISSUNIT;
*SUBPOPN (variable1) = (value); *Only use subpopn statement if needed;
```

```
WEIGHT VISWT; *or replace with your normalized weight if required for your analysis;
CLASS (variable2);
TABLES (variable2);
OUTPUT nsum wsum sewgt totper setot atlev1 atlev2 / FILENAME=[output dataset] REPLACE
TABLECELL=DEFAULT;
RUN;
```

In the above example, replace the parentheses with the information named in the parentheses. When health centers are missing a data element for all visits, those health centers' visits should be excluded from your analysis. If a subset of health centers' visits must be excluded due to complete missingness, replace VISWT with normalized version of the weight, and add a SUBPOPN statement to correctly subset to health centers' visits of interest. Refer to Section 7.2 above, for more guidance on weight normalization to account for complete missingness.

When using SAS-callable SUDAAN software, sort the input data set in the order specified in the NEST statement, in this case by sampling strata (STRATUM_S) followed by health center identifier (HCID_S) within STRATUM_S. If software other than SUDAAN is used to approximate the variances, other statements will be required by that software. The variance variables required by that software are the same as those include in the above example, which are further explained below in Section 7.3.1.

### Section 7.3.1 NEST Statement Variables

The SUDAAN NEST statement for variances at the visit-level is:

```
NEST STRATUM_S HCID_S / MISSUNIT;
```

Where:

STRATUM_S is the scrambled value of the original sampling stratum from which the health center was selected.

HCID_S is the scrambled identifier for the health center.

## Section 7.4 SAS Survey Procedures

The program below demonstrates how to calculate variance estimates using SAS SURVEYFREQ and SURVEYMEANS procedures:

For categorical variables:

```
PROC SURVEYFREQ DATA = (input data set);
TABLE VAR1; *Replace "VAR1" with the categorical variable of interest.
CLUSTER HCID_S;
STRATA STRATUM_S;
```

```
WEIGHT VISWT; *or replace with your normalized weight if required for your analysis;
ODS OUTPUT ONEWAY=(name of output);
RUN;
```

For continuous variables:

```
PROC SURVEYMEANS DATA = (input data set);
VAR VAR1; *Replace "VAR1" with the continuous variable of interest.
CLUSTER HCID_S;
STRATA STRATUM_S;
WEIGHT VISWT;
ODS OUTPUT STATISTICS=(name of output);
RUN;
```

In the above example, replace the parentheses with the information named in the parentheses. When health centers are missing a data element for all visits, those health centers' visits should be excluded from your analysis. If a subset of health centers' visits must be excluded due to complete missingness, replace VISWT with normalized version of the weight, and add a DOMAIN statement to correctly subset to health centers' visits of interest. Refer to Section 7.2 above, for more guidance on weight normalization to account for complete missingness.

## Section 7.5 R Survey Procedures

The R package "survey" can be utilized for complex survey analysis (https://cran.r-project.org/web/packages/survey/index.html). The R programs below demonstrate how to install the survey package, produce visit level weighted estimates, and calculate variance estimates.

```
install.packages("survey")
library(survey)
install.packages("tidyverse")
library(tidyverse)

#Using the "survey" package:
{variable name} <- svydesign(
        ids = ~ HCID_S,
        strata = ~ STRATUM_S,
        weights = ~ VISWT,
        data={input data frame})
```

Note: Replace curly brackets {} with the information named in the parenthesis

### Section 7.6 Stata Survey Procedures

The Stata programs below demonstrate how to use visit weights and calculate variance estimates with the svyset command (https://www.stata.com/manuals/svysvyset.pdf)

For categorical variables:

```
/*Set survey design*/
svyset HCID_S [pweight = VISWT], strata(STRATUM_S)

/*Specify one-way tables, change "VAR1" to categorical variable of interest*/
tab VAR1
svy: tab VAR1, count se
svy: tab VAR1, percent
```

For continuous variables:

```
/*Set survey design*/
svyset HCID_S, [pweight= VISWT], strata(STRATUM_S)

/*Specify one-way tables, change "VAR1" to continuous variable of interest*/
svy: mean VAR1
```

## Section 8 Survey Content

For the 2022 NAMCS HC Component public use data file, 77 variables were included; 60 (77.9%) variables include information on medical diagnoses, 8 (10.4%) variables include patient demographic information, 2 (2.6%) data items include visit information, and 7 (9.1%) variables include weights or other survey-related information.

Please refer to the 2022 NAMCS HC Component public use data file codebook for detailed information on the variables, including variable names, variable type, variable descriptions, and variable values. Additionally, unweighted frequencies for selected variables on the public use data file are available in Appendix A.

### Section 8.1 Demographic Item Missingness Rate

In the 2022 NAMCS HC Component public use data file, four (5.2%) demographic variables had an unweighted missingness rate that was greater than 5% including RACE, ETHNICITY, RACERETH and MARITAL.

The variables in the table below had an unweighted item missingness percentage greater than 5%. As explained in Section 7.2, some health centers contained complete missingness in certain variables. In

Table 8.1, two denominators are presented to demonstrate missingness. First, is the percent missing among all visits in all health centers (N=64) in the public use datafile. The second denominator is the percent missing among all visits in all health centers that do not have complete missingness in the diagnosis variable (N=54 or N=53).

**Table 8.1 Percent missing (unweighted) for demographic variables in the NAMCS HC Component public use data file with a missingness greater than 5%**

| Variable Name | Variable Description | % Missing (all visits) | % Missing[1] |
|---|---|---|---|
| RACERETH | Patient Race and Ethnicity | 12.04 | 4.13[3] |
| ETHNICITY | Patient Hispanic Ethnicity | 12.93 | 6.58[2] |
| RACE | Patient Race | 24.52 | 17.73[3] |
| MARITAL | Marital Status | 20.92 | 15.15[2] |

[1]Denominators vary as percentages exclude health centers with complete missingness.
[2]N=54 health centers.
[3]N=53 health centers.

## Section 8.2 Diagnosis Item Missingness Rate

In the 2022 NAMCS HC Component public use data file, 60 diagnosis variables (77.9%) had an unweighted missingness rate that was greater than 5%. It is expected that most of the diagnosis variables after the first-listed diagnosis variable will have a high missingness percentage as not all visits are expected to have multiple diagnoses.

The variables in the table below had an unweighted item missingness percentage greater than 5%. As explained in Section 7.2, some health centers contained complete missingness in certain variables. In Table 8.2, two denominators are presented to demonstrate missingness. First, is the percent missing among all visits in all health centers (N=64) in the public use data file. The second denominator is the percent missing among all visits in all health centers that do not have complete missingness in the diagnosis variable (N=58).

**Table 8.2 Percent missing (unweighted) for diagnoses variables in the NAMCS HC Component public use data file with a missingness greater than 5%**

| Variable Name | Variable Description | % Missing (All visits) | % Missing[1] |
|---|---|---|---|
| DX1 | Diagnosis #1 (ICD-10-CM), diagnosis code | 38.05 | 25.56 |
| DX2 | Diagnosis #2 (ICD-10-CM), diagnosis code | 56.99 | 48.32 |
| DX3 | Diagnosis #3 (ICD-10-CM), diagnosis code | 68.42 | 62.05 |
| DX4 | Diagnosis #4 (ICD-10-CM), diagnosis code | 76.22 | 71.43 |
| DX5 | Diagnosis #5 (ICD-10-CM), diagnosis code | 81.87 | 78.22 |

| | | | |
|---|---|---|---|
| DX6 | Diagnosis #6 (ICD-10-CM), diagnosis code | 86.37 | 83.63 |
| DX7 | Diagnosis #7 (ICD-10-CM), diagnosis code | 89.43 | 87.30 |
| DX8 | Diagnosis #8 (ICD-10-CM), diagnosis code | 91.58 | 89.88 |
| DX9 | Diagnosis #9 (ICD-10-CM), diagnosis code | 93.19 | 91.81 |
| DX10 | Diagnosis #10 (ICD-10-CM), diagnosis code | 94.35 | 93.21 |
| DX11 | Diagnosis #11 (ICD-10-CM), diagnosis code | 95.26 | 94.31 |
| DX12 | Diagnosis #12 (ICD-10-CM), diagnosis code | 96.00 | 95.19 |
| DX13 | Diagnosis #13 (ICD-10-CM), diagnosis code | 96.55 | 95.85 |
| DX14 | Diagnosis #14 (ICD-10-CM), diagnosis code | 96.98 | 96.37 |
| DX15 | Diagnosis #15 (ICD-10-CM), diagnosis code | 97.33 | 96.79 |
| DX16 | Diagnosis #16 (ICD-10-CM), diagnosis code | 97.62 | 97.14 |
| DX17 | Diagnosis #17 (ICD-10-CM), diagnosis code | 97.85 | 97.42 |
| DX18 | Diagnosis #18 (ICD-10-CM), diagnosis code | 98.08 | 97.69 |
| DX19 | Diagnosis #19 (ICD-10-CM), diagnosis code | 98.27 | 97.92 |
| DX20 | Diagnosis #20 (ICD-10-CM), diagnosis code | 98.42 | 98.10 |
| DX21 | Diagnosis #21 (ICD-10-CM), diagnosis code | 98.55 | 98.26 |
| DX22 | Diagnosis #22 (ICD-10-CM), diagnosis code | 98.68 | 98.42 |
| DX23 | Diagnosis #23 (ICD-10-CM), diagnosis code | 98.79 | 98.54 |
| DX24 | Diagnosis #24 (ICD-10-CM), diagnosis code | 98.89 | 98.67 |
| DX25 | Diagnosis #25 (ICD-10-CM), diagnosis code | 98.98 | 98.77 |
| DX26 | Diagnosis #26 (ICD-10-CM), diagnosis code | 99.06 | 98.87 |
| DX27 | Diagnosis #27 (ICD-10-CM), diagnosis code | 99.13 | 98.95 |
| DX28 | Diagnosis #28 (ICD-10-CM), diagnosis code | 99.19 | 99.02 |
| DX29 | Diagnosis #29 (ICD-10-CM), diagnosis code | 99.25 | 99.10 |
| DX30 | Diagnosis #30 (ICD-10-CM), diagnosis code | 99.30 | 99.16 |
| DX_TYPE1 | Diagnosis Type #1. Corresponds to Diagnosis #1 | 52.84 | 43.34 |
| DX_TYPE2 | Diagnosis Type #2. Corresponds to Diagnosis #2 | 68.28 | 61.89 |
| DX_TYPE3 | Diagnosis Type #3. Corresponds to Diagnosis #3 | 76.90 | 72.25 |
| DX_TYPE4 | Diagnosis Type #4. Corresponds to Diagnosis #4 | 82.66 | 79.17 |
| DX_TYPE5 | Diagnosis Type #5. Corresponds to Diagnosis #5 | 86.82 | 84.16 |
| DX_TYPE6 | Diagnosis Type #6. Corresponds to Diagnosis #6 | 89.91 | 87.87 |
| DX_TYPE7 | Diagnosis Type #7. Corresponds to Diagnosis #7 | 92.27 | 90.72 |
| DX_TYPE8 | Diagnosis Type #8. Corresponds to Diagnosis #8 | 93.97 | 92.75 |
| DX_TYPE9 | Diagnosis Type #9. Corresponds to Diagnosis #9 | 95.23 | 94.26 |
| DX_TYPE10 | Diagnosis Type #10. Corresponds to Diagnosis #10 | 96.14 | 95.36 |
| DX_TYPE11 | Diagnosis Type #11. Corresponds to Diagnosis #11 | 96.84 | 96.21 |
| DX_TYPE12 | Diagnosis Type #12. Corresponds to Diagnosis #12 | 97.41 | 96.88 |
| DX_TYPE13 | Diagnosis Type #13. Corresponds to Diagnosis #13 | 97.81 | 97.37 |
| DX_TYPE14 | Diagnosis Type #14. Corresponds to Diagnosis #14 | 98.14 | 97.77 |
| DX_TYPE15 | Diagnosis Type #15. Corresponds to Diagnosis #15 | 98.42 | 98.10 |
| DX_TYPE16 | Diagnosis Type #16. Corresponds to Diagnosis #16 | 98.65 | 98.38 |
| DX_TYPE17 | Diagnosis Type #17. Corresponds to Diagnosis #17 | 98.83 | 98.59 |

| | | | |
|---|---|---|---|
| DX_TYPE18 | Diagnosis Type #18. Corresponds to Diagnosis #18 | 98.99 | 98.79 |
| DX_TYPE19 | Diagnosis Type #19. Corresponds to Diagnosis #19 | 99.14 | 98.96 |
| DX_TYPE20 | Diagnosis Type #20. Corresponds to Diagnosis #20 | 99.24 | 99.09 |
| DX_TYPE21 | Diagnosis Type #21. Corresponds to Diagnosis #21 | 99.33 | 99.20 |
| DX_TYPE22 | Diagnosis Type #22. Corresponds to Diagnosis #22 | 99.41 | 99.30 |
| DX_TYPE23 | Diagnosis Type #23. Corresponds to Diagnosis #23 | 99.48 | 99.37 |
| DX_TYPE24 | Diagnosis Type #24. Corresponds to Diagnosis #24 | 99.54 | 99.45 |
| DX_TYPE25 | Diagnosis Type #25. Corresponds to Diagnosis #25 | 99.59 | 99.50 |
| DX_TYPE26 | Diagnosis Type #26. Corresponds to Diagnosis #26 | 99.63 | 99.56 |
| DX_TYPE27 | Diagnosis Type #27. Corresponds to Diagnosis #27 | 99.67 | 99.60 |
| DX_TYPE28 | Diagnosis Type #28. Corresponds to Diagnosis #28 | 99.70 | 99.64 |
| DX_TYPE29 | Diagnosis Type #29. Corresponds to Diagnosis #29 | 99.74 | 99.68 |
| DX_TYPE30 | Diagnosis Type #30. Corresponds to Diagnosis #30 | 99.76 | 99.71 |

[1]Denominators exclude health centers with complete missingness for all diagnosis variables (N=58 health centers).

# Section 9 Data Comparison

## Section 9.1 Public Use Data Files and Restricted Use Data File

Of the 64 participating health centers that were included in the 2022 NAMCS HC Component restricted use data file, all 64 (100%) were selected to create the public use data file sample. The 2022 public use data file contains 282,017 health center visits, for a weighted total of 109,087,913 health center visits (standard error of 19,896,515 health center visits). The 2022 NAMCS HC Component restricted use data file contains unweighted data from the same 64 health centers that submitted 5,640,370 health center visits, for a weighted total of 109,088,618 health center visits (standard error of 19,896,579 health center visits). A comparison of weighted frequencies for health center visits in the public use data file and restricted use data file is presented in Table 9.1.

**Table 9.1 Comparison of frequencies for health center visits on the public use data file (weighted n=109,087,913) and restricted use data file (weighted n=109,088,618) for NAMCS HC Component, 2022**

| Variable | Public Use Data file | | | | Restricted Use Data File | | | |
|---|---|---|---|---|---|---|---|---|
| | | Weighted | | | | Weighted | | |
| | Unweighted Count | Count | Std. Error | % | Unweighted Count | Count | Std. Error | % |
| **Age (in years)** | | | | | | | | |
| Under 1 | 6,024 | 2,288,302 | 411,698 | 2.1 | 122,534 | 2,308,440 | 412,527 | 2.1 |
| 1-17 years | 45,064 | 17,211,377 | 2,923,236 | 15.8 | 905,192 | 17,253,666 | 2,930,383 | 15.8 |
| 18-44 years | 99,228 | 38,421,536 | 6,790,823 | 35.2 | 1,975,554 | 38,304,878 | 6,780,308 | 35.1 |
| 45-64 years | 85,517 | 33,488,962 | 6,413,249 | 30.7 | 1,710,991 | 33,399,450 | 6,414,163 | 30.6 |
| 65-74 years | 31,066 | 12,025,463 | 2,740,872 | 11.0 | 620,942 | 12,081,788 | 2,733,533 | 11.1 |
| 75 years and over | 15,116 | 5,651,169 | 1,330,611 | 5.2 | 305,157 | 5,739,747 | 1,333,869 | 5.3 |
| Missing | <5 | 1,104 | 1,104 | 0.0 | 28 | 649 | 372 | 0.0 |
| **Sex** | | | | | | | | |
| Male | 105,699 | 41,028,536 | 7,177,835 | 37.6 | 2,109,098 | 40,899,996 | 7,149,913 | 37.5 |
| Female | 176,044 | 67,937,046 | 12,722,296 | 62.3 | 3,526,088 | 68,072,537 | 12,752,804 | 62.4 |
| Missing | 274 | 122,332 | 58,309 | 0.1 | 5,184 | 116,085 | 53,868 | 0.1 |
| **Visit month** | | | | | | | | |
| January | 24,403 | 9,637,222 | 1,869,987 | 8.8 | 488,037 | 9,637,161 | 1,869,843 | 8.8 |
| February | 21,269 | 8,241,409 | 1,480,435 | 7.6 | 425,460 | 8,242,383 | 1,480,504 | 7.6 |
| March | 23,017 | 9,133,987 | 1,749,641 | 8.4 | 460,333 | 9,133,975 | 1,749,420 | 8.4 |
| April | 22,177 | 8,627,374 | 1,644,153 | 7.9 | 443,632 | 8,628,944 | 1,644,429 | 7.9 |
| May | 22,267 | 8,655,310 | 1,672,140 | 7.9 | 445,243 | 8,652,907 | 1,672,003 | 7.9 |
| June | 23,038 | 8,871,436 | 1,658,288 | 8.1 | 460,800 | 8,873,100 | 1,658,351 | 8.1 |
| July | 21,794 | 8,487,764 | 1,601,614 | 7.8 | 435,793 | 8,485,961 | 1,601,677 | 7.8 |
| August | 25,695 | 9,795,541 | 1,818,491 | 9.0 | 513,933 | 9,795,745 | 1,818,375 | 9.0 |
| September | 25,793 | 9,840,784 | 1,783,007 | 9.0 | 515,871 | 9,841,231 | 1,783,090 | 9.0 |
| October | 24,277 | 9,494,903 | 1,754,538 | 8.7 | 485,482 | 9,493,768 | 1,754,615 | 8.7 |
| November | 24,980 | 9,379,282 | 1,669,960 | 8.6 | 499,563 | 9,493,768 | 1,669,787 | 8.6 |
| December | 23,307 | 8,922,901 | 1,666,742 | 8.2 | 466,223 | 8,924,503 | 1,666,919 | 8.2 |
| **Race** | | | | | | | | |
| AIAN | 2,279 | 881,481 | 188,471 | 0.8 | 46,054 | 894,900 | 186,278 | 0.8 |
| Asian | 8,209 | 3,896,949 | 1,574,629 | 3.6 | 163,968 | 3,877,577 | 1,563,474 | 3.6 |
| Black | 44,688 | 15,481,630 | 4,944,862 | 14.2 | 891,426 | 15,454,283 | 4,962,336 | 14.2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NHOPI | 1,952 | 964,033 | 529,511 | 0.9 | 41,100 | 1,010,010 | 557,207 | 0.9 |
| White | 143,411 | 54,129,306 | 9,687,712 | 49.6 | 2,869,416 | 54,192,231 | 9,698,795 | 49.7 |
| Other | 12,336 | 4,683,057 | 1,234,131 | 4.3 | 245,558 | 4,642,998 | 1,222,285 | 4.3 |
| Missing[1] | 69,142 | 29,051,458 | 10,088,550 | 26.6 | 1,382,848 | 29,016,620 | 10,032,231 | 26.6 |
| **Ethnicity** | | | | | | | | |
| Hispanic or Latino | 104,395 | 42,902,358 | 13,120,788 | 39.3 | 2,086,383 | 42,877,971 | 13,137,274 | 39.3 |
| Not Hispanic or Latino | 141,149 | 52,650,996 | 9,344,459 | 48.3 | 2,822,521 | 52,618,547 | 9,336,253 | 48.2 |
| Missing[1] | 36,473 | 13,534,559 | 2,646,004 | 12.4 | 731,466 | 13,592,101 | 2,648,372 | 12.5 |

[1]All health centers, including the health centers with complete missingness in race and ethnicity were included in the comparison of frequencies for race and ethnicity. When presenting analysis, data users should follow the normalization guidance provided in Section 7.2.

NOTE: All estimates provided in this table do not round to the nearest thousandth for comparison purposes. Data users should round to the nearest thousandth when presenting analyses as indicated in the presentation standards in Section 6.

# Section 10 Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA) Checklist for the 2022 NAMCS HC Component Public Use Data File

Table 10.1 below provides a Preferred Reporting Items for Complex Survey Analysis (PRICSSA) checklist (Seidenberg, Moser, & West, 2023) for users of the 2022 NAMCS HC Component public use data file. This information may be helpful to users when analyzing the 2022 NAMCS HC Component public use data file.

### 10.1 Preferred Reporting Items for Complex Sample Survey Analysis

| Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA) | Description |
|---|---|
| **Name of survey** | National Ambulatory Medical Care Survey Health Center Component |
| **Data collection mode** | EHR data submission |
| **Target population** | Federally qualified health centers (FQHCs) and FQHC look-alikes in the 50 U.S. states and the District of Columbia that used an EHR system |
| **Populations excluded** | Health Centers excluded:<br>- Indian Health Service Program facilities<br>- Did not provide healthcare services to the general U.S. population, or only provided care to special institutionalized populations such in prisons, nursing homes, homeless shelters, etc.<br>- Only provided dental services<br>- Were located on a military installation or outside of the 50 U.S. states and the District of Columbia |
| **Sample design** | Stratified systematic sampling |
| **Variance and standard error estimation** | PSU (HCID_S) and Stratum (STRATUM_S) variables were applied and Taylor Series Linearization was used to produce design-adjusted standard errors. |
| **Weighting** | VISWT, POPVST |
| **Presentation standards** | Proportions or percentages:<br>https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf<br>Rates and counts:<br>https://www.cdc.gov/nchs/data/series/sr_02/sr02_202.pdf |
| **Unweighted total sample size** | 282,017 visits |
| **Weighted total sample size** | 109,087,913 visits |
| **Response rate (unweighted)** | 25.1% |
| **Location of example code** | See Section 7 |

# Section 11 Research Data Center

NCHS operates the Research Data Center (RDC) to allow researchers access to restricted-use data. The RDC is responsible for protecting the confidentiality of survey respondents, study subjects, and institutions while providing access to restricted-use data for statistical purposes. The 2022 NAMCS HC Component restricted use data file, which contains unmasked and additional data from all visits at participating health centers (N=5,640,370 visits), can be accessed through the Federal and NCHS RDC. In addition, the 2022 NAMCS HC Component restricted use data file will be linked to other vital and administrative records such as the National Death Index (NDI), U.S. Housing and Urban Development (HUD) administrative data, and others. The linked data will both expand the analytic utility of the NAMCS HC Component data and provide the opportunity to conduct a vast array of studies focused on the associations between a variety of health factors, health care utilization, housing situations, and mortality.

For information on how to access the 2022 NAMCS HC Component restricted use data file through the RDC, please see: https://www.cdc.gov/rdc/b1datatype/Dt1224a.htm.

# Section 12 References

Health Level Seven International. HL7 CDA® R2 Implementation Guide: National Health Care Surveys (NHCS), R1 STU Release 3.1 - US Realm. Available at: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=385. Accessed November 29, 2023.

Lumley T. "survey: analysis of complex survey samples." R package version 4.2. 2023. Available at: https://cran.r-project.org/web/packages/survey/index.html.

Parker JD, Talih M, Malec DJ, et al. National Center for Health Statistics data presentation standards for proportions. National Center for Health Statistics. Vital Health Stat 2(175). 2017. Available at: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf

Parker JD, Talih M, Irimata KE, Zhang G, Branum AM, Davis D et al. National Center for Health Statistics data presentation standards for rates and counts. National Center for Health Statistics. Vital Health Stat 2(200). 2023. DOI: https://dx.doi.org/10.15620/cdc:124368.

Seidenberg AB, Moser RP, West BT. Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA). J Surv Stat Methodol 11(4). 2023. https://doi.org/10.1093/jssam/smac040.

StataCorp. Stata 18 Survey Data Reference Manual. College Station, TX: Stata Press. 2023. Available at: https://www.stata.com/manuals/svysvyset.pdf.

Williams SN, Ukaigwe J, Ward BW, Okeyode T, Shimizu IM. Sampling procedures for the collection of electronic health record data from federally qualified health centers, 2021-2022 National Ambulatory Medical Care Survey. National Center for Health Statistics. Vital Health Stat Series 2(203). 2023. DOI: https://dx.doi.org/10.15620/cdc:127730.

# Appendix A Unweighted frequencies for health center visits

**Appendix Table A.1. Unweighted frequencies for health center visits on the public use data file, National Ambulatory Medical Care Survey Health Center Component, 2022 (n=282,017)**

| Variable | Description | Count | % |
|---|---|---|---|
| **YEAR** | Survey year | | |
| 2022 | | 282,017 | 100 |
| **DAY** | Day of the week | | |
| 1 | Sunday | 1,697 | 0.6 |
| 2 | Monday | 54,076 | 19.2 |
| 3 | Tuesday | 60,480 | 21.5 |
| 4 | Wednesday | 58,219 | 20.6 |
| 5 | Thursday | 55,532 | 19.7 |
| 6 | Friday | 47,745 | 16.9 |
| 7 | Saturday | 4,268 | 1.5 |
| **MONTH** | Month of visit | | |
| 1 | January | 24,403 | 8.7 |
| 2 | February | 21,269 | 7.5 |
| 3 | March | 23,017 | 8.2 |
| 4 | April | 22,177 | 7.9 |
| 5 | May | 22,267 | 7.9 |
| 6 | June | 23,038 | 8.2 |
| 7 | July | 21,794 | 7.7 |
| 8 | August | 25,695 | 9.1 |
| 9 | September | 25,793 | 9.2 |
| 10 | October | 24,277 | 8.6 |
| 11 | November | 24,980 | 8.9 |
| 12 | December | 23,307 | 8.3 |
| **MARITAL** | Marital status | | |
| -9 | Missing | 59,010 | 20.9 |
| D | Divorced | 14,717 | 5.2 |
| L | Legally Separated | 3,567 | 1.3 |
| M | Married | 68,275 | 24.2 |
| O | Other | 23 | 0.0 |
| S | Separated | 46,783 | 16.6 |
| T | Domestic Partner | 3,700 | 1.3 |
| U | Unmarried | 77,297 | 27.4 |
| W | Widowed | 8,645 | 3.1 |
| **AGE_GROUP** | Patient age group | | |
| -9 | Missing | 2 | 0 |
| 1 | Less than 18 years | 51,088 | 18.1 |
| 2 | 18-44 years | 99,228 | 35.2 |
| 3 | 45-64 years | 85,517 | 30.3 |
| 4 | 65 years or more | 46,182 | 16.4 |
| **ETHNICITY** | Patient Hispanic ethnicity | | |
| -9 | Missing | 36,473 | 12.9 |
| 1 | Hispanic or Latino | 104,395 | 37.0 |
| 2 | Not Hispanic or Latino | 141,149 | 50.1 |
| **RACE** | Patient race | | |
| -9 | Missing | 69,142 | 24.5 |
| 1 | AIAN | 2,279 | 0.8 |

| | | | |
|---|---|---|---|
| 2 | Asian | 8,209 | 2.9 |
| 3 | Black | 44,688 | 15.9 |
| 4 | NHOPI | 1,952 | 0.7 |
| 5 | White | 143,411 | 50.9 |
| 6 | Other | 12,336 | 4.4 |
| **RACERETH** | Patient race and Hispanic ethnicity | | |
| -9 | Missing | 33,963 | 12.0 |
| 1 | White | 86,824 | 30.8 |
| 2 | Black | 42,807 | 15.2 |
| 3 | Hispanic | 103,447 | 36.7 |
| 4 | Other | 14,976 | 5.3 |
| **SEX** | Patient sex | | |
| -9 | Missing | 274 | 0.1 |
| 1 | Male | 105,699 | 37.5 |
| 2 | Female | 176,044 | 62.4 |