

Sampling Procedures for the Collection of Electronic Health Record Data From Federally Qualified Health Centers, 2021–2022 National Ambulatory Medical Care Survey

Data Evaluation and Methods Research



Centers for Disease Control and Prevention
National Center for Health Statistics

NCHS reports can be downloaded from:
<https://www.cdc.gov/nchs/products/index.htm>.

Copyright information

All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

Suggested citation

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>.

For sale by the U.S. Government Publishing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20401–0001
Printed on acid-free paper.

NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics

Series 2, Number 203

June 2023

Sampling Procedures for the Collection of Electronic Health Record Data From Federally Qualified Health Centers, 2021–2022 National Ambulatory Medical Care Survey

Data Evaluation and Methods Research

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
June 2023

National Center for Health Statistics

Brian C. Moyer, Ph.D., *Director*

Amy M. Branum, Ph.D., *Associate Director for Science*

Division of Health Care Statistics

Carol J. DeFrances, Ph.D., *Director*

Alexander Strashny, Ph.D., *Associate Director for Science*

Contents

- Abstract 1
- Background 1
- Sampling Methods and Procedures, 2021 and 2022 NAMCS HC Component 3
 - Sampling Frame Creation and Eligibility Requirements 3
 - Sampling Strata 4
- Distribution of Resulting HC Sample to Strata 4
- Summary 4
- References. 5
- Appendix I. Sampling Procedures for the 2021 National Ambulatory Medical Care Survey Health Center Component 6
- Appendix II. Sampling Procedures for the 2022 National Ambulatory Medical Care Survey Health Center Component 8

Text Figure

- In-person, manual data collection process for the 2012–2020 National Ambulatory Medical Care Survey
Community Health Center Component. 2

Text Table

- Percent distribution of eligible and ineligible health centers used for sample frame creation for
the 2021–2022 National Ambulatory Medical Care Survey Health Center Component 3

Sampling Procedures for the Collection of Electronic Health Record Data From Federally Qualified Health Centers, 2021–2022 National Ambulatory Medical Care Survey

by Sonja N. Williams, M.P.H., Joy Ukaigwe, M.S., Brian W. Ward, Ph.D., Titilayo Okeyode, M.Sc., and Iris M. Shimizu, Ph.D.

Abstract

Background

As part of modernization efforts, in 2021 the National Ambulatory Medical Care Survey (NAMCS) began collecting electronic health records (EHRs) for ambulatory care visits in its Health Center (HC) Component. As a result, the National Center for Health Statistics (NCHS) needed to adjust the approaches used in the sampling design for the HC Component. This report provides details on these changes to the 2021–2022 NAMCS.

Sampling Methods and Procedures

For the 2021 and 2022 NAMCS HC Component sampling frame, NCHS received a listing of all federally qualified health centers (FQHCs) and FQHC look-alikes in the United States in 2020 from the Health Resources and Services Administration (HRSA). An FQHC is an HC that receives funding from HRSA to provide services to people who are medically underserved. An FQHC look-alike is similar but does not receive funding. This listing included 1,463 HCs, of which 1,400 were eligible for the sampling frame. A stratified (organized information into groups based on certain criteria) random sample of 50

FQHCs and FQHC look-alikes was drawn as the primary sample, along with a reserve (or backup) sample of 100 HCs.

Results

For the 2021 sample, 95.7% of HCs in the source files from which the sampling frame was compiled were eligible for selection based on eligibility criteria. Among the 4.3% of ineligible HCs, 88.9% were ineligible because they lacked an EHR system or did not provide healthcare services to the public. For the 2022 sample, 85.6% of HCs in the source file were eligible for selection based on the eligibility criteria and the removal of all HCs selected for the 2021 sample. Among the 14.4% of HCs not eligible for the 2022 sample, 70.0% were selected for the 2021 sample, and 26.8% were ineligible because they lacked an EHR system or did not provide healthcare services.

Keywords: advanced practice provider • community health center • EHR • federally qualified health center look-alikes • healthcare provider • sampling design • NAMCS

Background

The National Ambulatory Medical Care Survey (NAMCS), administered by the National Center for Health Statistics (NCHS), is a national survey designed to meet the need for objective, reliable information about the provision and use of ambulatory medical care services in the United States. First fielded in 1973, the survey started with the collection of data from a sample of visits to nonfederally employed office-based physicians who were primarily engaged in direct patient care. To broaden its ability to capture the provision and use of ambulatory care, NAMCS also began sampling health centers (HCs) in 2006. HCs are community-based and

patient-directed organizations that “deliver comprehensive and affordable primary healthcare services to the nation’s most vulnerable populations, including people experiencing homelessness, agricultural workers, residents of public housing, and veterans” (1). Federally qualified health centers (FQHCs) are HCs that receive funding from the Department of Health and Human Services’ Health Resources and Services Administration (HRSA) to provide these services. These HCs must meet specific requirements set by HRSA to receive this funding. FQHC look-alikes are HCs that meet all the requirements from HRSA but do not receive funding. This expansion to include HCs in NAMCS was important because FQHCs often serve as “safety net” care settings, where health

care is provided for populations that are underserved or live in areas underserved by healthcare resources, and they are significant contributors in the primary care space (2,3).

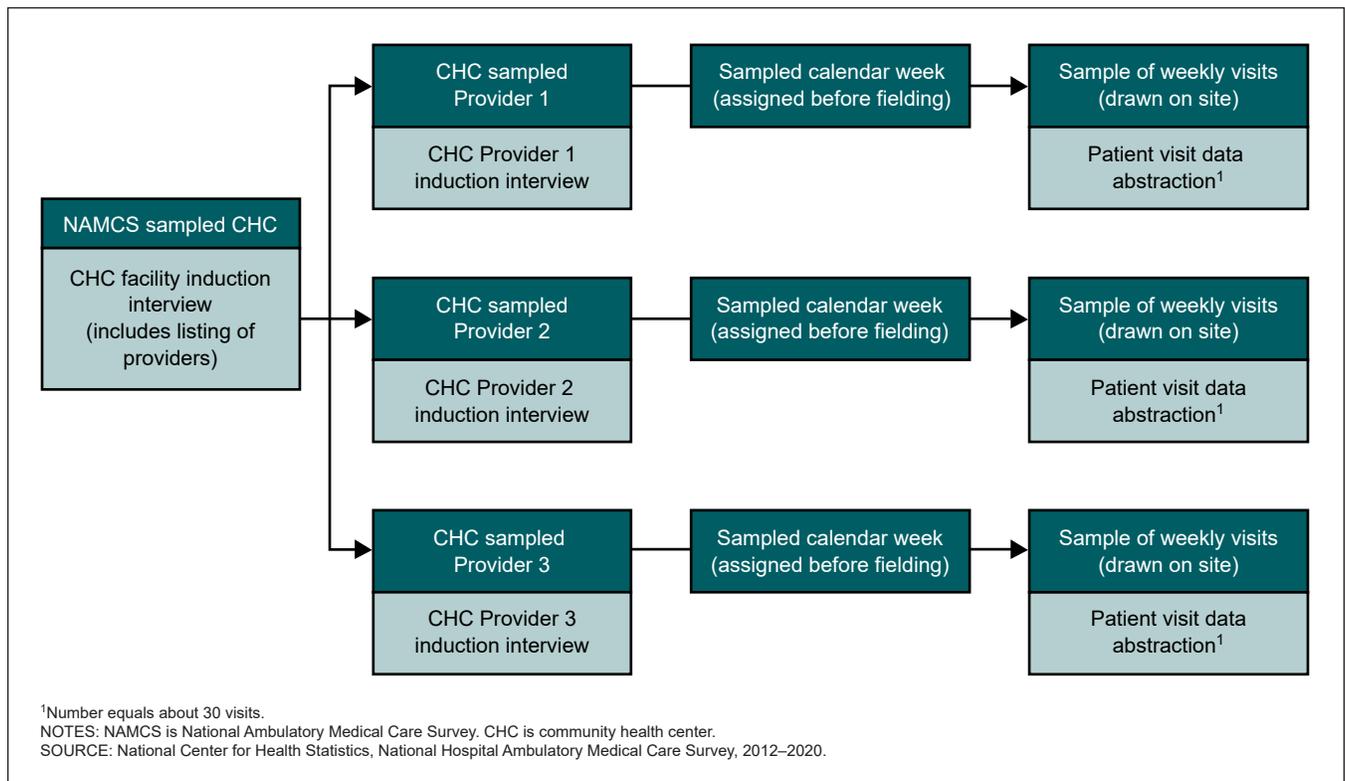
The sampling of HCs continued until 2012, when it was decided that a separate national sample of HC delivery sites should be drawn. This independent sample of HC delivery sites would allow for nationally representative estimates on HCs to be produced, providing more information on ambulatory care delivery at these locations. Before this change, HCs were only included within the traditional NAMCS physician sample, which limited the number of estimates that could be produced in that setting (4). This new sample of HC delivery sites included service sites of FQHCs and FQHC look-alikes and urban Indian Health Service outpatient clinics. Once a site was sampled, a facility interview was conducted with the HC site director or similar representative at that HC delivery site (Figure). This interview collected basic characteristics about the site, as well as a listing of all physicians, physician assistants, nurse practitioners, and certified nurse midwives who were currently providing direct patient care at the site. From this listing, up to three of these HC providers were randomly selected to receive a provider interview. Additionally, for each of these randomly selected providers, a random sample of approximately 30 patient visits seen by that provider was selected for which the visit records would be manually abstracted (physical review and gathering of information from medical records). These visits were selected from a week that was predetermined for the

sampled site during the NAMCS HC sampling process before fielding the survey. More details on this process can be found in the NAMCS data documentation (5).

NAMCS continued using this sampling approach and data collection procedures through 2020. However, in recent years NCHS leadership has desired to modernize NAMCS data collection procedures to better use electronic health records (EHRs) and other electronic data, increase the amount of data collected each year from survey respondents, decrease burden for survey respondents, and increase the timeliness and availability of the data collected. In 2019, NCHS, partnering with HRSA, committed to change its HC data collection procedures beginning with the 2021 NAMCS. This shift included three major changes. First, NAMCS would move from conducting manual in-person abstraction to collect patient visit data to collecting these data through HC submission of EHRs using the NCHS “HL7 CDA R2 Implementation Guide: National Health Care Surveys Release 1, DSTU Release 1.2—US Realm” (6). Second, instead of collecting data from individual delivery sites and selected HC providers, data would be collected from the entire FQHC or FQHC network. Finally, NAMCS would move from collecting only a sample of HC visits from a predetermined week to collecting data for the entire calendar year.

These combined changes created what is now called the NAMCS HC Component, and they are ultimately expected to increase the timeliness and analytic capabilities of NAMCS. However, as part of this change, new sampling procedures

Figure. In-person, manual data collection process for the 2012–2020 National Ambulatory Medical Care Survey Community Health Center Component



for the NAMCS HC Component were needed. As described previously, the previous NAMCS HC sample was based on manual identification of providers within the HC and then on sampling a small number of visits of the identified providers. One of the goals of the redesign was to minimize burden for the sampled HCs, and removing the need for identifying providers was key in achieving this goal. This report provides details on the procedures used to draw a sample of FQHCs and FQHC look-alikes for the 2021 and 2022 NAMCS.

Sampling Methods and Procedures, 2021 and 2022 NAMCS HC Component

Sampling Frame Creation and Eligibility Requirements

Beginning in 2021, the targeted universe (specific population targeted) for the NAMCS HC Component is FQHCs and FQHC look-alikes in the 50 U.S. states and the District of Columbia, which provide ambulatory (or direct outpatient) care to the public and use an EHR system in one or more of their delivery sites. Unlike previous years, Indian Health Service outpatient clinics were not included in the targeted universe in the 2021 and 2022 NAMCS HC Component because of complexities of the survey redesign and the changes to sampling procedures. Although these Indian Health Service outpatient clinics were not included in the 2021 and 2022 NAMCS, there is potential for drawing a complementary sample of these clinics for inclusion in future surveys after this new method of collecting HC data through EHR transmission has been established.

To create the sampling frame and draw the sample, NCHS worked with HRSA to use a nationally representative database that contains a list of all HCs in the United States. For the 2021 sample, this list included 1,463 HCs. To ensure

that only survey-eligible FQHCs and FQHC look-alikes were included, any HC that met the following conditions was omitted from the sampling frame (Table):

- HCs that did not have an EHR system
- HCs that did not provide healthcare services to the general U.S. population, such as those that exclusively serve institutional populations (for example, prisons, nursing homes or long-term care facilities, homeless shelters, etc.)
- HCs where only dental services were provided
- HCs located on a military installation or outside of the 50 U.S. states and the District of Columbia

Duplicate entries—HCs with the same name, address, and administrator contact information—were removed from the database. The resulting sampling frame of 1,400 HCs was used to draw the primary and backup samples for the 2021 NAMCS HC Component. Of the 63 FQHCs and FQHC look-alikes in the original database that were ineligible for the 2021 NAMCS HC Component sampling frame, 24 were omitted because they did not have an EHR system, 32 were omitted for not providing healthcare services to the public or only providing dental services, and 7 were omitted because they were located on a military installation or outside of the 50 U.S. states and the District of Columbia or appeared as duplicate listings.

To create a sampling frame for the 2022 NAMCS HC Component, the same database for 2021 was updated and used, which contained 1,482 HCs. Using financial support from the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) for fiscal year 2021 (7), the 2022 NAMCS HC Component sample was expanded to initially add 60 respondent HCs to the 50 respondent HCs from the 2021 sample, resulting in patient visit data being collected for 110 FQHCs and FQHC look-alikes during the 2022 NAMCS. For this expansion, 60 HCs (54 FQHCs ultimately fielded) were selected for the primary sample and 120 additional HCs for the backup sample. All the eligibility

Table. Percent distribution of eligible and ineligible health centers used for sample frame creation for the 2021–2022 National Ambulatory Medical Care Survey Health Center Component

Eligibility status	2021		2022	
	Number	Percent	Number	Percent
Total	1,463	100.0	1,482	100.0
Eligible	1,400	95.7	1,269	85.6
Ineligible				
No electronic health record system	24	1.7	26	1.8
Did not provide healthcare service to the general population or only provided dental services	32	2.2	31	2.1
Other ¹	7	0.5	7	0.6
Included in previous year's sample	149	10.1

... Category not applicable.

¹Includes health centers located on a military installation or outside of the 50 U.S. states and the District of Columbia, or that appeared as duplicate listings.

NOTE: The sampling frame was the total number of all eligible health centers for that year.

SOURCE: U.S. Department of Health and Human Services, Health Resources and Services Administration Electronic Handbooks Database.

criteria noted previously were applied and duplicate entries were removed. However, an additional step for the 2022 sampling process was that the 149 HCs sampled for 2021 (50 in the main sample, 99 in the backup sample, and 1 that was dropped from the 2021 database) were removed from the sampling frame that was used for selecting the sample added in 2022. For 2022, this process yielded a sampling frame of 1,269 eligible HCs. The 2022 NAMCS HC Component target respondent sample was reduced from 60 to 54 because of budget restraints. To accommodate the needed change, six randomly selected HCs were removed from the sample in four strata.

Sampling Strata

The sampling strata were defined by U.S. Census region (Northeast, Midwest, South, and West) and metropolitan statistical area (MSA) status (MSA and non-MSA). Within each stratum, HCs were randomly numbered to ensure randomness in the selected sample from each stratum. To minimize the risk of disclosing the identity of HCs in each stratum, this report provides summaries by range instead of single percentages.

For the 2021 NAMCS HC Component sampling frame, the distribution of HCs by geographical region ranged from 17.0% to 33.5%. By MSA status, 70.7% of HCs in the sampling frame were in an MSA. When combined, the regional and MSA distribution ranged from 3.2% to 21.8%.

The 2022 NAMCS HC Component sampling frame had a similar range in distribution, with 16.1% to 34.2% of HCs located in the different geographical regions. By MSA status, 72.8% of HCs in the sampling frame were in an MSA. When combined, the regional and MSA distribution ranged from 2.4% to 22.6%.

Within each sampling stratum, HCs were sorted by the nine U.S. Census divisions and ordered within each division by random numbers previously assigned to the HCs. From the randomly ordered HCs in each stratum, systematic random sampling was used to select two nonoverlapping samples. The size of the first, or primary sample, was the targeted number of participating sample HCs desired from the survey, while the size of the second, or backup sample, was twice the size of the primary sample. If any of the HCs in the primary sample from a stratum were ineligible or declined to participate in the NAMCS HC Component, HCs in the backup sample from the same stratum were contacted for participation as needed until the targeted number of sample stratum HCs agreed to submit their EHRs for patient visits. For 2021, the primary and backup samples included 50 and 100 HCs, respectively. For 2022, 54 (reduced from 60 after the sample was drawn) and 120 HCs were included in the primary and backup samples, respectively.

This method of using a backup sample had not been previously used with NAMCS; however, it was implemented

beginning in 2021 to increase the likelihood that NCHS would be able to collect visit data from 50 HCs in 2021, and an additional 54 HCs in 2022. The procedures used to draw the 2021 and 2022 NAMCS HC Component samples can be found in the technical appendixes. Appendix I provides details on the 2021 NAMCS HC Component sampling procedures. Appendix II provides details on the 2022 procedures.

Distribution of Resulting HC Sample to Strata

An ideal sample has an equitable distribution of HCs relative to the sampling frame across geographical areas. Small sample sizes for the 2021 and 2022 NAMCS resulted in differences between the population and sample distributions of HCs by U.S. region, MSA status, and the combination of region and MSA status (the sampling strata themselves). However, the goal of mirroring the distribution of the HC population by region and MSA status was achieved for each of the samples.

Regarding U.S. region and MSA status, the range of the difference in the distribution for the 2021 samples relative to the sampling frame was 3.7 to 7.0 percentage points. The 2021 samples had a higher percentage of HCs in MSAs (5.3 percentage point difference) compared with the sampling frame. The range of the distribution of HCs in the 2022 sample relative to the sampling frame ranged from 0.8 to 8.2 percentage points. The 2022 samples had similar percentages of HCs in MSAs (0.5 percentage point difference) to the sampling frame.

For the 2021 NAMCS HC Component sampling strata (defined by U.S. region and MSA status), all differences ranged from -5.9 to 4.3 percentage points. For the 2022 NAMCS, all differences ranged from -9.2 to 4.3 percentage points.

Summary

Although NAMCS has been the leading data collection tool for ambulatory health care across the United States for almost 50 years and has been collecting patient visit data from HCs since 2006, changes in the way healthcare information for patients is recorded have required a change in how NAMCS data are collected and how HCs are selected for participation. Beginning in 2021, NAMCS began gathering ambulatory healthcare data directly from a nationally representative sample of FQHCs and FQHC look-alikes. This modernization necessitated a new sampling approach for the 2021 and 2022 NAMCS HC Component. This method is expected to be used in future years, whenever the HC sample is updated or replaced.

Stratified (organized information into groups based on certain criteria) random samples of HCs were selected for the 2021 and 2022 NAMCS HC Component with strata defined by U.S. Census region and MSA status to ensure nationally representative estimates. Additionally, primary and backup

samples were used to enhance the ability to obtain targeted numbers of participants for both years. The 2021 sample was expanded in 2022 with support from the OS-PCORTF FY2021. This report of the NAMCS HC Component sampling methods is expected to give researchers a better understanding of the sampling design and (once collected) how to use and interpret these NAMCS HC Component data. Consequently, researchers will have a better understanding of ambulatory health care provided at FQHCs and FQHC look-alikes and can help inform patient care at these “safety net” providers (3) across the United States.

References

1. U.S. Department of Health and Human Services, Health Resources and Services Administration. What is a health center? 2021. Available from: <https://bphc.hrsa.gov/about-health-centers/what-health-center>.
2. Politzer RM, Yoon J, Shi L, Hughes RG, Regan J, Gaston MH. Inequality in America: The contribution of health centers in reducing and eliminating disparities in access to care. *Med Care Res Rev* 58(2):234–48. 2001.
3. Wakefield M. Federally qualified health centers and related primary care workforce issues. *JAMA* 325(12):1145–6. 2021.
4. National Center for Health Statistics. 2012 NAMCS CHC micro-data file documentation. 2015. Available from: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/doc2012_chc.pdf.
5. National Center for Health Statistics. 2014 NAMCS CHC micro-data file documentation. 2019. Available from: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/doc2014_chc-508.pdf.
6. Health Level Seven International (HL7). HL7 CDA R2 implementation guide: National Health Care Surveys (NHCS), release 1, DSTU release 1.2—US realm. 2021. Available from: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=385.
7. U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. Enhancing surveillance of maternal health clinical practices and outcomes with federally qualified health centers’ (FQHCs) electronic health records visit data. 2021. Available from: <https://aspe.hhs.gov/enhancing-surveillance-maternal-health-clinical-practices-outcomes-federally-qualified-health>.

Appendix I. Sampling Procedures for the 2021 National Ambulatory Medical Care Survey Health Center Component

This appendix provides instructions for selecting a stratified (organized information into groups based on certain criteria) list sample of health centers (HCs) for the 2021 National Ambulatory Medical Care Survey Health Center (NAMCS HC) Component with sampling strata defined by four U.S. Census regions (Northeast, Midwest, South, and West) and metropolitan statistical area (MSA) status (MSA and non-MSA). For 2021, a primary sample of 50 HCs was drawn, in addition to a backup sample of 100 HCs.

Common Notation Used for Sampling Specifications

Let

h stand for MSA status (MSA or non-MSA)

j stand for an HC

M, m stand for the total number of HCs in the sampling frame and sample, respectively

$I(B)$ stand for identity function

$$= \begin{cases} 1 & \text{if condition } B \text{ is met, and} \\ 0 & \text{if condition } B \text{ is not met} \end{cases}$$

Int stand for the sampling interval used in selecting a sample

r stand for U.S. Census region (Northeast, Midwest, South, and West)

Sampling Procedures

Step 1. Prepare the 2021 NAMCS sampling frame

Using a national database, only eligible federally qualified health centers (FQHCs) and FQHC look-alikes were included in the sample. Any HCs that met the following conditions were deemed ineligible and omitted from the sampling frame:

- HCs that did not have an electronic health record (EHR) system

- HCs that did not provide healthcare services to the general U.S. population, such as those that exclusively serve institutional populations (for example, prisons, nursing homes or long-term care facilities, homeless shelters, etc.)
- HCs where only dental services were provided
- HCs located on a military installation or outside of the 50 U.S. states and the District of Columbia

Additionally, duplicate entries in the U.S. Department of Health and Human Services' Health Resources and Services Administration electronic handbooks—HCs with the same name, address, and administrator contact information—were also removed. Then, each HC sampling stratum was defined by U.S. Census region r and MSA status h .

Step 2. Array (arrange) HCs within sampling strata

To array HCs within the sampling strata:

- Assign each HC a random number. This randomization may be done in the whole sampling frame or within each sampling stratum.
- Within each sampling stratum r, h (defined by U.S. Census region r and MSA status h), sort the HCs by U.S. Census division.
- Within U.S. Census division, order HCs by the random numbers assigned in step 2a.
- Serially number the randomly ordered HCs in each sampling stratum r, h .

Step 3. Select the 2021 NAMCS sample of HCs from each sampling stratum

The following instructions apply for the sample in each sampling stratum r, h :

- Calculate the sampling interval for selecting the primary sample for the r, h stratum (carrying two decimal places):

$$Int(HC)_{r,h} = M_{r,h} / mA_{r,h} \quad [3.1]$$

where

$M_{r,h}$ = Total number of HCs listed in the sampling frame for sampling stratum r,h [3.2]

$mA_{r,h}$ = Total number of HCs to be selected to the primary sample from sampling stratum r,h

$$= \begin{cases} 3 & \text{if } h = 2 \text{ (if not MSA)} \\ 9 & \text{if } h = 1 \text{ and } r = 1, 2, \text{ or } 4 \\ 11 & \text{if } h = 1 \text{ and } r = 3 \end{cases} \quad [3.3]$$

b. Select a random start $R_{r,h}$ between 0 and $Int(HC)_{r,h}$, inclusive where $Int(HC)_{r,h}$ is defined in equation [3.1].

c. Select to the primary sample from stratum r,h the first HC for which the serial number assigned in step 2d is the nearest integer greater than or equal to

$$R_{r,h} + Int(HC)_{r,h} \cdot k, \quad k = 0, 1, 2, 3, \dots$$

The total number of HCs selected from the sampling stratum r,h should be within 1 of the sample number targeted for that stratum according to equation [3.3]. That is, if $mA'_{r,h}$ is the number of HCs initially selected to the sample for stratum r , then $mA'_{r,h}$ should be in the range $mA_{r,h} \pm 1$.

d. Select a backup sample of HCs to use if any primary sample HC should be ineligible or a nonrespondent. The size of the backup sample from each stratum should be twice the size mA of the primary sample from that stratum to minimize the possibility that the backup sample will be depleted, or completely used, before a total of $mA_{r,h}$ in scope (HCs with all required criteria) and respondent HCs are inducted from stratum r,h .

First, remove the primary sample HCs from the frame in each sampling stratum to form the “remaining” sampling frame for the backup sample.

Let

$M2_{r,h} = M_{r,h} - mA_{r,h}$ stand for the number of HCs remaining in the frame for the r,h stratum after selection of the primary sample.

$$Int2(HC)_{r,h} = M2_{r,h} / [2 \cdot mA_{r,h}] \quad [3.4]$$

stands for the sampling interval for selecting the backup sample from the remaining frame for each r,h stratum, where $mA_{r,h}$ is defined in equation [3.3] above.

Repeat steps b and c with the reduced sampling frame and revised sampling interval $Int2(HC)_{r,h}$

e. For each contacted sample HC that is confirmed ineligible or nonrespondent (from either the primary or the backup sample), add to the sample an HC from the backup sample, adding those HCs to the sample in order of the serial numbers assigned to the HCs in step 2d.

f. After $m_{r,h}$ respondent HCs have been confirmed as a participant and inducted from sampling stratum r,h , insert

into the record for each contacted sample HCj (including all participants, nonrespondents, and ineligible HCs) that HC's sampling weight $\omega_{j,2021}$ where

$$\omega_{j,2021} = \frac{M_{r,h}}{m_{r,h}} \cdot I \left(\begin{matrix} j \text{ is in 2021 sample, in} \\ \text{region } r, \text{ and MSA status } h \end{matrix} \right)_j \quad [3.5]$$

$M_{r,h}$ = Total number (before any sampling) of HCs listed in the 2021 sampling frame for stratum r,h

$m_{r,h}$ = Total number of sample HCs that were contacted for participation in the 2021 HC survey, regardless of scope or response status determined for those HCs. This number includes all respondents, nonrespondents, and ineligible HCs among those contacted.

Note that according to the instructions above, the sample selected from each sampling stratum r,h is an equal probability sample. The sampling weight $\omega_{j,2021}$ is the inverse of the product of the probabilities of HCj being selected at some time k , given that HCj was not selected to the sample at times $k-1$ or earlier.

Appendix II. Sampling Procedures for the 2022 National Ambulatory Medical Care Survey Health Center Component

This appendix provides technical details for selecting a stratified (organized information into groups based on certain criteria) list sample of health centers (HCs) for the 2022 National Ambulatory Medical Care Survey Health Center (NAMCS HC) Component with sampling strata defined by four U.S. Census regions (Northeast, Midwest, South, and West) and metropolitan statistical area (MSA) status (MSA or non-MSA). For 2022, before it was reduced to 54, a primary sample of 60 HCs was drawn, in addition to a backup sample of 120 HCs. The HCs in the 2021 final total sample are also included in the 2022 final total sample, along with the 50 participants and all other HCs that were contacted for participation but were either ineligible or refusals (see Appendix I).

Common Notation Used for Sampling Specifications

Let

h stand for MSA status (MSA or non-MSA)

j stand for an HC

M, m stand for the total number of HCs in the sampling frame and sample, respectively

$I(B)$ stand for identity function

$$= \begin{cases} 1 & \text{if condition } B \text{ is met, and} \\ 0 & \text{if condition } B \text{ is not met} \end{cases}$$

Int stand for the sampling interval used in selecting a sample

r stand for U.S. Census region (Northeast, Midwest, South, and West)

Sampling Procedures

Step 1. Prepare the 2022 NAMCS sampling frame

Following the instructions given in Appendix I, step 1, compile the 2022 sampling frame from lists of HCs obtained from the Department of Health and Human Services' Health Resources Services Administration in 2021.

Step 2. Array (arrange) HC within sampling strata

Define the HC sampling strata by region r , and MSA status h . Also assign numbers to the HCs and array them within strata as instructed in Appendix I, step 2.

Step 3. Assign frame weights to HCs in the 2022 frame

Let

$M_{r,h,2022(HC)}$ = Total number of HCs listed in the 2022 sampling frame for sampling stratum defined by region r , and MSA status h , after completing step 1. These are HCs eligible for the total 2022 HC sampling frame (including those selected in the 2021 primary sample and in the 2021 backup sample).

$m_{r,h(2021HCsamp)}$ = Number of all HCs in the 2022 sampling frame for region r and MSA status h that are also in the 2021 primary sample or the 2021 backup sample, regardless of whether they were ever contacted to participate in the survey. This number includes HCs in the 2021 primary and backup samples that:

- Participate in the survey
- Were contacted but refused to participate in 2021
- Were contacted but confirmed ineligible for the survey
- Were never contacted and asked to participate

a. Obtain counts, $M_{r,h,2022(HC)}$ and $m_{r,h(2021HCsamp)}$ for each sampling stratum defined by region r and MSA status h .

b. After obtaining the two counts in the previous step, delete all $m_{r,h(2021HCsamp)}$ HCs from the 2022 sampling frame. The frame remainder is referred to in the following as the sampling list from which HCs will be selected for adding to the 2022 sample.

c. In the record for each HC remaining in the frame, insert the frame weight for the HC. That weight for HCs in region r and MSA h is:

$$FW_{r,h,2022} = M_{r,h,2022(HC)} / \left[\frac{M_{r,h,2022(HC)} - m_{r,h(2021HCsamp)}}{m_{r,h(2021HCsamp)}} \right] \quad [4.1]$$

Step 4. Select the HCs to be added to the 2022 HC sample

The following instructions apply for the sample in each sampling stratum r,h .

- a. Calculate the sampling interval for selecting the initial or primary sample for the stratum r,h (carrying two decimal places):

$$Int(HC)_{r,h,2022s} = M_{r,h,2022(HC)} / mA_{r,h,2022} \quad [4.2]$$

where

$M_{r,h,2022(HC)}$ = Total number of HCs listed in the 2022 sampling lists for sampling stratum defined by region r and MSA status h

$mA_{r,h,2022}$ = Total number of HCs to be selected from sampling stratum r,h from the 2022 sampling list

$$= \begin{cases} 4 & \text{if } h = 2 \text{ (if not MSA)} \\ 9 & \text{if } h = 1 \text{ and } r = 1, 2, \text{ or } 4 \\ 17 & \text{if } h = 1 \text{ and } r = 3 \end{cases} \quad [4.3]$$

- b. Select a random start $R_{r,h,2022}$ between 0 and $Int(HC)_{r,h,2022}$, inclusive where $Int(HC)_{r,h,2022}$ is defined in equation [4.2].
- c. In the order arrayed in step 2, accumulate the frame weights of HCs remaining in the 2022 HC list and assign to each HC the cumulative sum that results after adding the weight for that HC to the sum.

Then select to the sample the first HC whose cumulative sum of frame weights is greater than or equal to

$$R_{r,h,2022} + Int(HC)_{r,h,2022} \cdot k, \quad k = 0, 1, 2, 3, \dots$$

where $R_{r,h,2022}$ is defined in step b above. That is, if $mA'_{r,h,2022}$ is the number of HCs initially added to the 2022 sample for stratum r,h then $mA'_{r,h,2022}$ should be in the range $mA_{r,h,2022} \pm 1$.

- d. For the 2022 added sample, select a backup sample of HCs to use if any added primary sample HC should be ineligible or a nonrespondent.

First remove the primary sample HCs selected from the 2022 sampling list in each sampling stratum to form the “remaining” sampling list for the 2022 backup sample.

Let

$$Int2(HC)_{r,h,2022} = B_{r,h,2022} / [2 \cdot mA_{r,h,2022}]$$

Denote the sampling interval for selecting the backup sample from the remaining sampling list for each r,h stratum

where

$B_{r,h,2022}$ = Total of frame weights summed across all HCs remaining in the stratum defined by region r and MSA status h , after removing the primary sample selected from the 2022 sampling list.

$mA_{r,h,2022}$ is defined in equation [4.3] above.

Repeat steps b and c with the reduced sampling list and revised sampling interval $Int2(HC)_{r,h,2022}$

- e. For each contacted sample HC from stratum r,h that is confirmed ineligible or nonrespondent, add one HC from the backup sample from the same r,h stratum to the sample. Add those HCs in the order of serial numbers assigned to the HCs in the 2022 sampling list in step 2.
- f. After $m_{r,h,2022}$ HCs have been confirmed participants from sampling stratum r,h , insert into the record for each contacted sample HC j (including all participants, nonrespondents, and ineligible HCs) that HC’s sampling weight $\omega_{j,2022}$

where

$$\omega_{j,2022} = FW_{r,h} \left(\frac{Int(HC)_{r,h}}{FW_{r,h}} \right) \cdot I \left(\begin{array}{l} j \text{ is in sample added} \\ \text{for 2022, in region } r, \\ \text{and MSA status } h \end{array} \right)_j \quad [4.4]$$

$$= Int(HC)_{r,h} \cdot I(j \text{ is in sample added for 2022 in region } r \text{ and MSA status } h)_j \quad [4.5]$$

For the backup sample HCs that are contacted, their sampling weights would be the same as in equation [4.5], with $Int(HC)_{r,h}$ the same as for selecting the primary sample, given that the contacted backup sample HC shares the same weight as the primary sample HC.

Vital and Health Statistics Series Descriptions

Active Series

- Series 1. Programs and Collection Procedures**
Reports describe the programs and data systems of the National Center for Health Statistics, and the data collection and survey methods used. Series 1 reports also include definitions, survey design, estimation, and other material necessary for understanding and analyzing the data.
- Series 2. Data Evaluation and Methods Research**
Reports present new statistical methodology including experimental tests of new survey methods, studies of vital and health statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Reports also include comparison of U.S. methodology with those of other countries.
- Series 3. Analytical and Epidemiological Studies**
Reports present data analyses, epidemiological studies, and descriptive statistics based on national surveys and data systems. As of 2015, Series 3 includes reports that would have previously been published in Series 5, 10–15, and 20–23.

Discontinued Series

- Series 4. Documents and Committee Reports**
Reports contain findings of major committees concerned with vital and health statistics and documents. The last Series 4 report was published in 2002; these are now included in Series 2 or another appropriate series.
- Series 5. International Vital and Health Statistics Reports**
Reports present analytical and descriptive comparisons of U.S. vital and health statistics with those of other countries. The last Series 5 report was published in 2003; these are now included in Series 3 or another appropriate series.
- Series 6. Cognition and Survey Measurement**
Reports use methods of cognitive science to design, evaluate, and test survey instruments. The last Series 6 report was published in 1999; these are now included in Series 2.
- Series 10. Data From the National Health Interview Survey**
Reports present statistics on illness; accidental injuries; disability; use of hospital, medical, dental, and other services; and other health-related topics. As of 2015, these are included in Series 3.
- Series 11. Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey**
Reports present 1) estimates of the medically defined prevalence of specific diseases in the United States and the distribution of the population with respect to physical, physiological, and psychological characteristics and 2) analysis of relationships among the various measurements. As of 2015, these are included in Series 3.
- Series 12. Data From the Institutionalized Population Surveys**
The last Series 12 report was published in 1974; these reports were included in Series 13, and as of 2015 are in Series 3.
- Series 13. Data From the National Health Care Survey**
Reports present statistics on health resources and use of health care resources based on data collected from health care providers and provider records. As of 2015, these reports are included in Series 3.

- Series 14. Data on Health Resources: Manpower and Facilities**
The last Series 14 report was published in 1989; these reports were included in Series 13, and are now included in Series 3.
- Series 15. Data From Special Surveys**
Reports contain statistics on health and health-related topics from surveys that are not a part of the continuing data systems of the National Center for Health Statistics. The last Series 15 report was published in 2002; these reports are now included in Series 3.
- Series 16. Compilations of Advance Data From Vital and Health Statistics**
The last Series 16 report was published in 1996. All reports are available online; compilations are no longer needed.
- Series 20. Data on Mortality**
Reports include analyses by cause of death and demographic variables, and geographic and trend analyses. The last Series 20 report was published in 2007; these reports are now included in Series 3.
- Series 21. Data on Natality, Marriage, and Divorce**
Reports include analyses by health and demographic variables, and geographic and trend analyses. The last Series 21 report was published in 2006; these reports are now included in Series 3.
- Series 22. Data From the National Mortality and Natality Surveys**
The last Series 22 report was published in 1973. Reports from sample surveys of vital records were included in Series 20 or 21, and are now included in Series 3.
- Series 23. Data From the National Survey of Family Growth**
Reports contain statistics on factors that affect birth rates, factors affecting the formation and dissolution of families, and behavior related to the risk of HIV and other sexually transmitted diseases. The last Series 23 report was published in 2011; these reports are now included in Series 3.
- Series 24. Compilations of Data on Natality, Mortality, Marriage, and Divorce**
The last Series 24 report was published in 1996. All reports are available online; compilations are no longer needed.

For answers to questions about this report or for a list of reports published in these series, contact:

Information Dissemination Staff
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782

Tel: 1–800–CDC–INFO (1–800–232–4636)

TTY: 1–888–232–6348

Internet: <https://www.cdc.gov/nchs>

Online request form: <https://www.cdc.gov/info>

For e-mail updates on NCHS publication releases, subscribe online at: <https://www.cdc.gov/nchs/email-updates.htm>.

**U.S. DEPARTMENT OF
HEALTH & HUMAN SERVICES**

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782-2064

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300

FIRST CLASS MAIL
POSTAGE & FEES PAID
CDC/NCHS
PERMIT NO. G-284



For more NCHS Series Reports, visit:
<https://www.cdc.gov/nchs/products/series.htm>