

Comparative Health System Performance Initiative: Compendium of U.S. Health Systems, 2022, Outpatient Site Linkage File, Technical Documentation

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract Number: 47QRAA18D00BQ/75Q80123F80002

Prepared by:

Mathematica, Washington, DC
Emily Hague
Miaomiao Shen
Katie Bisson
Genna Cohen
David Jones
Eugene Rich

AHRQ Publication No. 24-0061
July 2024



This page has been left blank for double-sided copying.

Contents

Acknowledgments.....	v
I. Introduction.....	1
II. Data Sources	3
III. Methodology.....	5
A. Identifying Outpatient Sites.....	5
B. Linking Outpatient Sites with Health Systems.....	5
C. Identifying the Census Tract and County Subdivision for Each Outpatient Site	6
D. Using Federal Data Sources to Describe Outpatient Sites’ Locations	7
IV. Outpatient Site Linkage File Contents.....	11
V. Caveats and Limitations	15
A. Definition of a Health System	15
B. Reliance on IQVIA OneKey Data	15
C. Comparing Related Variables in the Outpatient Site Linkage File and Systems List.....	16
D. Differences Between the Outpatient Site Linkage File and Group Practice Linkage File.....	17
E. Limitations of RUCA Data.....	18
Appendix A. Data Dictionary – Outpatient Site Linkage File.....	19
Appendix B. Rural-Urban Commuting Area Code Descriptions	21

Tables

IV.1. Outpatient sites in the Compendium’s outpatient site linkage file.....	11
IV.2. Local area characteristics of outpatient sites	12
V.1. Compendium systems with no affiliated sites in the outpatient site linkage file.....	16
V.2. Differences between the outpatient site linkage file and group practice linkage file	17

This page has been left blank for double-sided copying.

Acknowledgments

We would like to acknowledge a number of organizations and individuals who provided support and guidance during the development of the Compendium of U.S. Health Systems, 2022, Outpatient Site Linkage File:

- Agency for Healthcare Research and Quality: Dan Miller, Leeann Comfort, Jay Crosson, Lingrui Liu, Mike Furukawa, Bruce Seeman, and Doreen Bonnett
- Data vendor: IQVIA™
- Mathematica staff: Andrew McGuirk, Sandi Nelson, Eric Dehus, Michael Barna, Neetu Jain, and Linda Molinari

This page has been left blank for double-sided copying.

I. Introduction

In 2015, the Agency for Healthcare Research and Quality (AHRQ) created the Comparative Health System Performance (CHSP) Initiative to study how healthcare systems promote evidence-based practices in delivering care.ⁱ AHRQ’s goal is to understand the factors that affect health systems’ use of patient-centered outcomes research (PCOR) and identify best practices in disseminating and using PCOR.

AHRQ supports the ongoing work of the CHSP Initiative by providing the Compendium of U.S. Health Systems, a list of health systems in the United States with details about their structure, staffing, and program participation. The Compendium is available on the AHRQ website: <https://www.ahrq.gov/chsp/index.html>.

As part of the CHSP Initiative, AHRQ and Mathematica developed a series of publicly available data resources for researchers, policymakers, and other stakeholders who want to understand how health systems can improve the value of healthcare. These data resources include the 2016, 2018, 2020, 2021, and 2022 Compendium of U.S. Health Systems and hospital linkage files; 2016 and 2018 group practice linkage files; and now the 2022 outpatient site linkage file.

AHRQ developed the health system data files (referred to as the systems list in this document) using information from several data sources that identify systems and their members. In addition to the names and locations, the systems list includes characteristics such as the number of physicians, hospitals, and nursing homes in the system and the number of hospital discharges from system hospitals.

AHRQ also created a file linking health systems with their outpatient sites (referred to as the “outpatient site linkage file” in this document). The outpatient site linkage file includes information on outpatient sites where care is delivered by a physician, nurse practitioner, or physician assistant, such as site name, street address, city, state, and ZIP Code. In addition, the file includes the health system name and Compendium health system ID for outpatient sites linked to health systems in the systems list. The file also includes variables describing characteristics of the geographic divisions in which the outpatient site is located, such as information about healthcare shortage designations and the site’s rurality.

The outpatient site linkage file is publicly available so users can identify outpatient sites within health systems and then link to other data sources to examine aspects of their performance, such as cost and quality of care.

This document summarizes the approach taken to create the 2022 outpatient site linkage file. In Section II, we summarize the data sources used to create the file. Section III describes the methodology used to create and refine the file. In Section IV, we describe the variables contained in the file. Finally, Section V describes caveats that should be considered when using the outpatient site linkage file.

ⁱ Additional information on the CHSP Initiative can be found at <https://www.ahrq.gov/chsp/index.html>.

This page has been left blank for double-sided copying.

II. Data Sources

We use the IQVIA OneKey data to identify outpatient sites in the United States that have a physician, nurse practitioner, or physician assistant, and their links to health systems in the systems list. The OneKey data include frequent updates of information on health systems, physicians, advanced practice clinicians, and healthcare facilities nationwide, including hospitals, clinical practice sites, nursing homes, and home health care organizations.

The data come from a combination of telephone surveys and administrative sources. They contain system- and facility-level identifying information such as name, location, and unique OneKey identifiers to link records over time.

Most importantly for the purposes of creating the outpatient site linkage file, the OneKey data also describe relationships between healthcare providers and other organizations via ownership, management, leasing, purchasing, and contracting mechanisms. In particular, the data indicate relationships between health systems included in the systems list and outpatient sites. We use this information to identify the links between the outpatient sites and health systems to create the outpatient site linkage file.ⁱⁱ

In addition to the OneKey data used to identify outpatient sites with a physician, nurse practitioner, or physician assistant and link them to systems, we used data from federal sources to characterize the census tract of each outpatient site. We used data from the Health Resources and Service Administration (HRSA) to identify sites in a medically underserved area (MUA), an area with a medically underserved population (MUP), or a health professional shortage area (HPSA).ⁱⁱⁱ We also used data from the U.S. Department of Agriculture's Economic Research Service to identify each site's rural-urban commuting area (RUCA) code.^{iv}

We merged the outpatient sites with these federal data using each site's geographic coordinates. We describe these variables added to the outpatient site linkage file in Section IV.

ⁱⁱ An overview of the OneKey data collection methods is available in the Compendium of U.S. Health Systems, 2020, Technical Documentation at <https://www.ahrq.gov/chsp/data-resources/compendium.html>.

ⁱⁱⁱ More information on these HRSA designations is available at <https://bhw.hrsa.gov/workforce-shortage-areas/shortage-designation>.

^{iv} More information on RUCAs is available at <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/>.

This page has been left blank for double-sided copying.

III. Methodology

The unrestricted OneKey data contain 490,030 outpatient sites. Here, we first describe our approach to identify outpatient sites with a physician, nurse practitioner, or physician assistant. We then describe our approach to link outpatient sites with systems, and finally, to link outpatient sites with other federal data sources.

A. Identifying Outpatient Sites

We first limit the records in the OneKey data to outpatient sites. The most common types of outpatient sites before applying any restrictions are medical group practice sites, independent physician practices, health clinics, physical medicine and rehabilitation centers, outpatient surgical centers, and imaging centers.

Medical groups are outpatient healthcare centers that provide general or specialized services to patients. Typically, they include two or more prescribers (e.g., physicians, nurse practitioners, physician assistants) and one or more non-prescribers (e.g., nurses, medical technicians). Independent physician practices are outpatient sites with a single physician. The unrestricted data also include less numerous outpatient sites, such as health departments, medical spas, and alternative medicine clinics.

We then limit the 490,030 outpatient sites from the OneKey data to 485,759 sites in the 50 U.S. states and the District of Columbia (DC). Finally, we further limit the sites to those with a confirmed physician (medical doctor or doctor of osteopathic medicine), nurse practitioner, or physician assistant, resulting in a set of 283,138 outpatient sites in the file.

B. Linking Outpatient Sites With Health Systems

The OneKey data identify relationships between outpatient sites and corporate parents where applicable. Health systems (referred to as integrated delivery networks [IDNs] in the OneKey data) are one type of corporate parent identified in the data, and each health system has a unique OneKey ID.

Because we created the systems list using the OneKey data, it also includes this unique OneKey ID. Thus, we link the outpatient sites and their health systems to the systems list using the unique OneKey IDs. We also report the name of the corporate parent owner and the entity type (corporate owner, IDN, or blank for outpatient sites that do not have an owner identified in the OneKey data).

For outpatient sites with owners that appear in the Compendium systems list, we report the Compendium health system ID and other information from the systems list. Some smaller, geographically defined health systems (subsystems) are nested within larger systems (parent systems). In the systems list, we aggregate the information for subsystems under their parent systems and only report the information for the parent systems. Similarly, in the outpatient site linkage file, we report the parent system information for outpatient sites linked to subsystems.

After linking outpatient sites to systems and outpatient sites in subsystems to their parent systems, we have 69,637 outpatient sites (24.6% of sites in the outpatient site linkage file) linked

to systems identified in the Compendium of U.S. Health Systems, 2022. The count of outpatient sites in the outpatient site linkage file differs from the count of outpatient centers in the systems list. That is because the systems list enumerates all outpatient centers linked to the system (including, for example, sites staffed by physical and occupational therapists), and the outpatient site linkage file enumerates only sites with at least one physician or advanced practice clinician.

C. Identifying the Census Tract and County Subdivision for Each Outpatient Site

We identify the locations of the outpatient sites at the most granular geographic level possible so that we can link the sites with information about those geographic divisions (see Section III.D). The OneKey data include latitude and longitude (`geo_addr_lat_nbr` and `geo_addr_lon_nbr`), as well as the field `geo_lvl_cd`, which indicates the method used to assign the latitude and longitude coordinates.

The variable `geo_lvl_cd` reports three levels of precision for coordinates:

- RT (interpolated rooftop, information at street-front level),
- Z9 (centroid of a nine-digit ZIP Code tabulation area), and
- Z5 (centroid of a five-digit ZIP Code tabulation area).

For a handful of observations, `geo_lvl_cd` is na (address did not match) or `geo_lvl_cd` is not populated and the coordinates are missing. Both RT- and Z9-designated coordinates are highly accurate to the street level, as nine-digit ZIP Codes represent a segment or one side of a street.^v In contrast, because five-digit ZIP Codes are more general and cover post office facility or delivery areas, they often overlap with multiple geographic divisions.

We join the coordinates for each outpatient site with RT or Z9 coordinates (which represent a point on the map) with spatial files that represent the boundaries of the geographic divisions where they are located.^{vi} We use this method to identify the census tract for each outpatient site and to construct the outpatient site linkage variable `tract`. This variable `tract` represents the full state-county-census tract 11-digit Federal Information Processing Standards (FIPS) code.

The first two digits of the variable `tract` represent the state FIPS code, the next three digits represent the county FIPS code, and the final six digits represent the census tract. For example, for census FIPS 01001020100, the first 01 represents the state of Alabama, the next 001 represents Autauga County, and the final 020100 represents the specific census tract located in Autauga County, Alabama.

We use the same process to identify each outpatient site's county subdivision. County subdivisions are portions of a county that represent a town, township, or community area.^{vii} We

^v More information on the definition and specificity of ZIP Codes is available at <https://faq.usps.com/s/article/ZIP-Code-The-Basics>.

^{vi} We performed a point-in-polygon spatial join, using R's `sf` package, to merge information from the U.S. Census Bureau's 2022 TIGER/line shapefiles.

^{vii} See <https://www2.census.gov/geo/pdfs/reference/GARM/Ch8GARM.pdf> for more detail about county subdivisions.

use this process to construct the outpatient site linkage file variable `county_sub`, which represents the state, the county, and the county-subdivision 10-digit FIPS code.

The first two digits of the variable `county_sub` represent the state FIPS code, the next three digits represent the county FIPS code, and the final five digits represent the county subdivision. For example, for county subdivision FIPS 0100190171, the first 01 represents the state of Alabama, the following 001 represents Autauga County, and the final 90171 represents the county subdivision of Autaugaville.

Because five-digit ZIP Codes cover much larger areas that often overlap with multiple census tracts and county subdivisions and because the coordinates reflect the centroids of these large areas, we cannot directly attribute the outpatient sites with Z5 or missing coordinates to census tracts or subdivisions. Instead, for these sites, we use the U.S. Department of Housing and Urban Development's USPS ZIP Code Crosswalk Files to map the site's addresses to census tracts or county subdivisions.

If a site is associated with a five-digit ZIP Code that maps to a single census tract, then the census tract is populated using the crosswalk. Likewise, if a site is associated with a five-digit ZIP Code that maps to a single county subdivision, then the county subdivision field is populated using the crosswalk. In addition, when nine-digit ZIP Codes are available for sites with Z5 coordinates, we use an internally developed crosswalk to link the site to the census tract or county subdivision corresponding to that nine-digit ZIP Code.

If multiple census tracts or county subdivisions are associated with a site and its five-digit ZIP Code and we also cannot link the site to a census tract or county subdivision based on its nine-digit ZIP Code, then we leave the census tract or county subdivision field blank.^{viii}

D. Using Federal Data Sources To Describe Outpatient Sites' Locations

We use the census tract and county subdivision FIPS codes to merge additional information from other federal sources about the geography where each outpatient site was located. We use the census tract FIPS code to identify the primary and secondary RUCA codes for each outpatient site. RUCA codes are a method of classifying rurality based on population density, urbanization, and daily commuting information. The primary RUCA code is based on the area's largest commuting flows, and the secondary RUCA code further subdivides the RUCA codes based on secondary commuting flows.

Because the most recent available RUCA codes at the time of publication were based on 2010 census tracts, we convert each site's census tract to its 2010 equivalent.^{ix} For outpatient sites that

^{viii} Connecticut counties were replaced in June 2022 with county-equivalent planning regions. We identify the December 2022 county-equivalent planning region for sites in Connecticut using the first five digits of the census tract or county subdivision FIPS code, as described earlier in this section. Therefore, outpatient sites in Connecticut that are missing census tract or county subdivision information will also have missing data in the county field.

^{ix} Census tract equivalency data to convert 2020 census tract identifiers to 2010 census tract identifiers are available at https://www2.census.gov/geo/docs/maps-data/data/rel2020/tract/tab20_tract20_tract10_natl.txt. RUCA is missing in the file for sites when we cannot identify a 2010 census tract equivalent.

have a nine-digit ZIP Code but no 2020 census tract or identifiable 2010 census tract equivalent, we use an internally developed crosswalk to link the site directly to the 2010 census tract corresponding to that nine-digit ZIP Code. We then use the 2010 equivalent census tracts to merge to the RUCA codes for each outpatient site.

We also use the census tract, county subdivision, and county FIPS codes to identify outpatient sites located in a healthcare shortage area (MUA, MUP, or HPSA). HRSA oversees MUA, MUP, and HPSA designations. MUAs designate primary care shortage areas for a specific area, and MUPs designate shortages for a specific population within an area that might face economic, cultural, or language barriers to healthcare.^x

HPSAs are another way to identify healthcare shortages, and, similar to MUAs and MUPs, a HPSA can be designated for an entire geographic area or for a specific population in an area.^{xi,xii} We identify outpatient sites as being in a shortage area if they are within a census tract, county subdivision, county, or some combination of these identified as a shortage area by HRSA with the designation for that shortage area active on December 31, 2022.^{xiii, xiv}

For outpatient sites within a healthcare shortage area, we also merge in unique identifiers for the shortage areas. The shortage area unique identifiers can be merged with the HRSA data to obtain additional information about the shortage areas, such as:

- Shortage area name,
- Index of Medical Underservice score (for MUAs and MUPs) or degree of shortage (for HPSAs),
- Information about which populations have designated shortages, and
- Relevant population data.

^x Examples of specific populations identified by MUPs include people experiencing homelessness, people with low income, people eligible for Medicaid, American Indian/Alaska Native people, and migrant farm workers.

^{xi} Examples of specific populations identified by HPSAs include low-income populations, homeless populations, and migrant farmworker populations.

^{xii} HPSAs can also be designated at the single point or facility level. We do not identify HPSAs at this level because the shortage variables are intended to describe the availability of healthcare resources in the geographic area surrounding outpatient sites. The facility-level designations might indicate that a specific site has a shortage of providers or that a site serves a population or geographic area with a shortage of providers.

^{xiii} Shortage area data can be downloaded from <https://data.hrsa.gov/data/download>. While shortage areas should not be designated at more than one geographic level (census tract, county subdivision, or county) and geographic areas should not be designated as both an MUA and MUP, we observe several such overlaps in data downloaded in February 2024. For example, CONCENTRA MEDICAL CENTER (compendium_os_id OS00100234) is associated with two MUAs (00840 and 00835) and an MUP (00801). This overlap occurs in densely populated areas because the HRSA data allows boundary overlaps of up to 2 square miles. We include all observed designations in the outpatient site linkage file.

^{xiv} We consider designations active if the Status Code for the shortage area is either “Designated” or “Proposed Withdrawal” on or before a designation date of December 31, 2022. We do not consider shortage areas active for the outpatient site linkage file if they are designated as “Withdrawn” as of that date or if they have a designation date on or after January 1, 2023.

Finally, we use information about each outpatient site's Core Based Statistical Area (CBSA) provided in the OneKey data to determine whether the site was in a metropolitan statistical area (an urban area of 50,000 or more inhabitants) or micropolitan statistical area (an area with at least 10,000 and up to 50,000 inhabitants).^{xv}

Outpatient sites located in neither a metropolitan nor micropolitan statistical area do not have an associated CBSA. CBSA information is used to develop an indicator for rurality. We consider an outpatient site to be rural if the site is not located in a metropolitan statistical area, consistent with the definition used by the Office of Management and Budget.^{xvi}

^{xv} We find that the CBSA provided by OneKey is out of date for 19 unique CBSAs. For the outpatient sites in those 19 CBSAs, we use a metro area history file available from the U.S. Census Bureau (<https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/historical-delineation-files.html>) to identify the outpatient sites' current CBSAs, replacing the outdated CBSAs in the data.

^{xvi} See <https://www.hhs.gov/guidance/document/defining-rural-population> for more information about rural definitions commonly used by federal agencies.

This page has been left blank for double-sided copying.

IV. Outpatient Site Linkage File Contents

The outpatient site linkage file contains 283,138 outpatient sites (Table IV.1), 69,637 of which are linked to Compendium health systems. The remaining 213,501 outpatient sites are not linked to a Compendium health system. Medical groups constitute 228,823 of the 283,138 outpatient sites. The remaining 54,315 outpatient sites are other types of outpatient facilities, such as independent physician practices, imaging centers, and outpatient surgical centers.

Table IV.1. Outpatient sites in the Compendium’s outpatient site linkage file

Outpatient Site Group	N
Total outpatient sites in Compendium file	283,138
Linked to Compendium health systems	69,637
Medical groups	63,329
Not linked to Compendium health systems	213,501
Medical groups	165,494

The outpatient site linkage file contains 33 variables, including each outpatient site’s:

- Name;
- Location (address, city, state, and ZIP Code);
- Unique outpatient site ID assigned as part of the development of the Compendium;
- Facility type (such as medical group, independent physician practice, or imaging center);
- Facility specialty (nested under facility type, with examples of medical group types including primary care, orthopedics, and cardiology);
- Count of physicians from OneKey; and
- If applicable, the site’s affiliated Compendium health system ID, name, and location.

We detail variables describing site geographic designations and characteristics below.

The outpatient site linkage file contains two linking variables. The OneKey ID (`onekey_id`) can be used to link outpatient sites to OneKey data. The unique health system identifier (`health_sys_id`) can be used to link outpatient sites to the systems in the Compendium systems list.

The outpatient site linkage file also contains variables that identify the corporate parents of outpatient sites. These variables include the corporate parent name (`corp_parent_name`), corporate parent OneKey ID (`corp_parent_id`), and corporate parent type (`corp_parent_type`). Two types of corporate parents are noted: health systems (IDNs) and corporate owners.^{xvii} When an outpatient site has corporate parent information reported in the linkage file but does not report a Compendium health system, it means the corporate parent is a corporate owner or an IDN that was determined to not meet the definition of a health system for the Compendium.

^{xvii} According to OneKey documentation, a corporate owner is “an organization that owns, leases, manages, or establishes strictly a purchasing affiliation with two or more healthcare delivery sites, but does not offer a continuum of healthcare. This is because the organization does not include at least one acute care hospital and one nonacute organization.”

Section V.A includes the definition of a health system for the Compendium and summarizes the exclusion criteria used to identify systems meeting this definition. More detailed information on how we use the definition to identify Compendium systems is available in the Compendium of U.S. Health Systems, 2022, Technical Documentation at <https://www.ahrq.gov/chsp/data-resources/compendium.html>.

Differences between (nonmissing) information provided for the corporate parent and Compendium health system information may exist because:

1. The corporate parent identified in the OneKey data is determined to be a subsystem of a Compendium health system. In these cases, the corporate parent information reflects the subsystem, and the Compendium system information reflects the parent system.
2. The corporate parent name and the Compendium health system name have slight differences. In these cases, the corporate parent information reflects the health system name from the OneKey data. The Compendium health system name reflects a slight cleaning of the system name, or it is based on other data sources used to develop the systems list.

In addition, we added variables created using federal data sources that describe the characteristics of the local areas where the outpatient sites are located, including the following:

- Indicators of healthcare shortage areas
- Indicators of rurality

Table IV.2 summarizes these variables.

Table IV.2. Local area characteristics of outpatient sites

Characteristics	Definition
Medically Underserved Area (MUA)	<p>This binary indicator shows whether the outpatient site is located in an MUA. An outpatient site is within an MUA if it is within a geographic area (census tract, county subdivision, or county) identified as an MUA by HRSA with the status of that designation active at the end of 2022. HRSA designates MUAs for geographic areas that lack access to primary care services.</p> <p>MUAs are identified based on the Index of Medical Underservice, which includes the ratio of providers to every 1,000 people, the percentage of the population below the federal poverty line, the percentage of the population age 65 and older, and the infant mortality rate.^a This variable takes the value 1 if the outpatient site is in an MUA, 0 if the outpatient site is not in an MUA, and missing if we cannot identify whether the outpatient site is in an MUA based on missing geographic division data (see Section III.C).</p> <p>When the indicator value is 1, we also populate a field with the unique identifiers for the MUA where the outpatient site is located. Users can obtain more information about the MUA by merging this unique MUP identifier with the HRSA data.</p>

Characteristics	Definition
Medically Underserved Population (MUP)	<p>This binary indicator shows whether the outpatient site is located in an MUP. An outpatient site is within an MUP if it is within a geographic area (census tract, county subdivision, or county) identified as an MUP by HRSA with the status of that designation active at the end of 2022. HRSA designates MUPs for geographic areas that have a population-specific lack of access to primary care services. The identified populations might face economic, cultural, or language barriers to accessing primary care.^b</p> <p>MUPs are identified based on the Index of Medical Underservice for a specific population, which includes the ratio of providers to every 1,000 people, the percentage of the population below the federal poverty line, the percentage of the population age 65 and older, and the infant mortality rate.^a This variable takes the value 1 if the outpatient site is in an MUP, 0 if the outpatient site is not in an MUP, and missing if we cannot identify whether the outpatient site is in an MUP based on missing geographic division data (see Section III.C).</p> <p>When the indicator value is 1, we also populate a field with the unique identifiers for the MUP where the outpatient site is located. Users can obtain additional information about the MUP, including about the specific population for which the shortage is designated, by merging this unique MUP identifier with the HRSA data.</p>
Primary Care Health Professional Shortage Area (HPSA)	<p>This binary indicator shows whether the outpatient site is in an HPSA. An outpatient site is within a primary care HPSA if it is within a geographic area (census tract, county subdivision, or county) identified as a primary care HPSA by HRSA with the status of that designation active at the end of 2022. HRSA designates geographic HPSAs for geographic areas with a shortage of providers; these shortages might be for an entire area or for a specific population within an area.^c</p> <p>Primary care HPSAs are identified based on the geographic area's population-to-provider ratio, the percentage of the population below the federal poverty line, an infant health index, and travel time to the nearest source of care outside the HPSA designation. This variable takes the value 1 if the outpatient site is in an HPSA, 0 if the outpatient site is not in an HPSA, and missing if we cannot identify whether the outpatient site is in an HPSA based on missing geographic division data (see Section III.C).</p> <p>When the indicator value is 1, we also populate a field with the unique identifiers for the HPSA where the outpatient site is located. Users can obtain more information about the HPSA, including whether the HPSA is for an entire area or for a specific population, and about the specific population for which the shortage is designated, by merging this unique HPSA identifier with the HRSA data.</p>
Rural Urban Commuting Area (RUCA) code	<p>An outpatient site's RUCA code is based on the census tract in which it is located, mapped to that tract's 2010 equivalent. The most recent RUCA codes available as of February 2024 are based on data from the 2010 census and the 2006–2010 American Community Survey.</p> <p>RUCA codes identify metropolitan, micropolitan, small town, and rural commuting areas based on the size and direction of the largest commuting flows. Secondary RUCA codes provide more detail about secondary commuting flows.^{d,e} The linkage files contain separate variables for the primary and secondary RUCA codes.</p>
Rurality	<p>This binary indicator shows whether the outpatient site is in a rural area. An outpatient site is rural if the site is not part of a metropolitan statistical area, consistent with the definition used by the Office of Management and Budget.^f</p>

^a HRSA’s scoring criteria for MUAs, MUPs, and HPSAs are available at <https://bhwh.hrsa.gov/workforce-shortage-areas/shortage-designation/scoring>.

^b Examples of specific populations identified by MUPs include people experiencing homelessness, people with low income, people eligible for Medicaid, American Indian/Alaska Native people, and migrant farm workers.

^c Examples of specific populations identified by HPSAs include low-income populations, homeless populations, and migrant farmworker populations.

^d More information on RUCAs is available at <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/>.

^e See Appendix B for RUCA code descriptions.

^f See <https://www.hhs.gov/guidance/document/defining-rural-population> for more information about rural definitions commonly used by federal agencies.

The values for all variables are missing when the specific values are missing in the data, or when a census tract or county subdivision cannot be identified for the outpatient site (see Section III.C, Identifying the Census Tract and County Subdivision for Each Outpatient Site). Appendix A contains a data dictionary for all variables included in the outpatient site linkage file.

V. Caveats and Limitations

This release of the outpatient site linkage file enables users to link Compendium health systems with their outpatient sites. When using the outpatient site linkage file, users should bear in mind a few caveats and limitations to the current methods for assigning outpatient sites to health systems.

A. Definition of a Health System

The Compendium definition of a health system follows:

A health system includes at least one hospital and at least one group of physicians that provides comprehensive care (including primary and specialty care) who are connected with each other and with the hospital through common ownership or joint management.^{xviii}

This definition is further operationalized as requiring that a health system:

1. Include at least one nonfederal acute care hospital,
2. Include, in total, at least 50 physicians, and
3. Include at least 10 primary care physicians.

If an organization does not meet the definition of a Compendium system, then affiliated outpatient sites are not identified in the outpatient site linkage file as part of a Compendium system. In addition, the Compendium systems linked to outpatient sites in the linkage file are parent systems; relationships between outpatient sites and subsystems are not directly identified in the Compendium system variables. Relationships with organizations not considered Compendium health systems may be identified in the corporate parent variable, as discussed in Section IV.

B. Reliance on IQVIA OneKey Data

Three health systems in the Compendium systems list are not found in the OneKey data and therefore have missing data for outpatient sites (Table V.1). In addition, the outpatient site linkage file relies on the set of outpatient sites and relationships to health systems identified in the OneKey data. These may not capture all outpatient sites or all relationships between these sites and health systems nationwide.

^{xviii} Foundation models of health system organization are considered a form of joint management. Joint participation in an accountable care organization is not by itself indicative of joint management. In addition, “group” is not synonymous with a separately organized medical group. A hospital that employs community-based physicians who provide comprehensive care (but are not organized as a medical group) would be considered a health system.

Table V.1. Compendium systems with no affiliated sites in the outpatient site linkage file

Compendium System ID	System Name
HSI00000116	Blanchard Valley Health System
HSI00000715	Munson Healthcare
HSI00000487	Holy Family Memorial

C. Comparing Related Variables in the Outpatient Site Linkage File and Systems List

The outpatient site linkage file enumerates outpatient sites with a confirmed physician (medical doctor or doctor of osteopathic medicine), nurse practitioner, or physician assistant. This restriction is distinct from the approach taken to create the variable in the systems list that enumerates all outpatient centers linked to each system regardless of the clinicians present.^{xix} Users can aggregate the count of outpatient sites in the outpatient site linkage file to the system level by using the system information provided in the outpatient site linkage file, but the count will differ from the number of outpatient centers reported in the system file.

The counts calculated from the outpatient site linkage file will always be less than the counts reported in the system file because the two datasets start with the same underlying data, and they use the same approaches to counting outpatient sites. The exception is that the outpatient site linkage file restricts outpatient sites to those with a confirmed physician, nurse practitioner, or physician assistant.^{xx}

The systems list contains counts of physicians by system, and the counts of physicians in the outpatient site linkage file can be summed by Compendium health system, but there are several key differences in these system-level counts. First, the physician counts in the systems list reflect a combination of OneKey data and the American Hospital Association's Annual Survey Database, and the counts in the linkage file reflect only the OneKey data.^{xxi}

Second, the physician counts in the linkage file reflect relationships between physicians and health systems specifically through outpatient sites. The physician counts in the systems list reflect relationships between physicians and health systems through all facility types and are intended to include all physicians in health systems. Thus, the counts in the systems list are typically much larger than the aggregate counts at the system level summed using the outpatient linkage file.

^{xix} Outpatient center counts are reported in the `grp_cnt` variable in the system file.

^{xx} Systems are identified in the outpatient site linkage file via the variables `health_sys_id` or `health_sys_name`.

^{xxi} Physician counts are reported in the `total_mds` variable in the system file.

D. Differences Between the Outpatient Site Linkage File and Group Practice Linkage File

Similar to the outpatient site linkage file, the group practice linkage file enables users to link data on group practices to the systems list. But the outpatient site linkage file differs in four main ways from the group practice linkage file:

- The unit of observation is the outpatient site in the outpatient site linkage file and the physician group in the group practice linkage file.
- The methodology used to link outpatient sites and physician groups to systems differs between the two files.
- The two files have differing requirements for minimum number of physicians.
- The outpatient site linkage file is constructed using OneKey as the sole data source, and the group practice linkage file leverages data from multiple sources.

Because of these differences, which we summarize in Table V.2, the outpatient site linkage file and group practice linkage file do not link together neatly (for example, related outpatient sites do not always nest neatly under a group practice), and related outpatient sites and group practices could link to different health systems in some cases.

Table V.2. Differences between the outpatient site linkage file and group practice linkage file

Consideration	Outpatient Site Linkage File	Group Practice Linkage File
Unit of observation	Physical address	Physician group as identified by a Tax Identification Number
Method of linking to systems	Record in OneKey data	Reconciliation across four approaches to linking group practices to health systems
Minimum physicians	At least one physician, nurse practitioner, or physician assistant	At least two physicians in a group
Data source	OneKey	Medicare data on provider practice and specialty supplemented with Medicare Advantage encounter data; the Compendium hospital linkage file; Medicare Provider Enrollment, Chain, and Ownership System data; Medicare claims; and OneKey data

We elaborate on the unit of observation and the method of linking to systems below.

1. Unit of Observation

The outpatient site linkage file contains site- or location-level data on all outpatient sites identified in the OneKey data that have at least one physician, nurse practitioner, or physician assistant. Sites include independent physician practices, physician group practices that do not bill Medicare (such as pediatric physician offices and medical spas), and outpatient sites that are not physician group locations (such as health department sites and imaging centers). In contrast, the group practice linkage file contains group practice- or Tax Identification Number-level data on physician practices that bill to Medicare.

Group practices identified in the group practice linkage file deliver services at physician practice sites but not the other types of outpatient sites included in the outpatient site linkage file. Furthermore, many group practices deliver services at more than one outpatient site.

2. Method of Linking to Systems

Outpatient centers are linked to health systems in the outpatient site linkage file based solely on the health system identified in the OneKey data. For outpatient sites linked with health systems that appear on the systems list, we report the Compendium health system ID and other information from the systems list. For outpatient sites with a corporate parent identified that is not on the systems list, we only report the name of the corporate parent (as discussed in Section IV). Outpatient sites with missing corporate parent information do not have a corporate parent identified in the OneKey data.

In contrast, the group practice linkage file uses a combination of four approaches to link group practices to health systems.^{xxii} Group practices may be linked to systems using the:

1. Tax Identification Number's CMS Certification number.
2. Level of hospital-based billing among their physicians.
3. Links between their organizational National Provider Identifier and health systems reported in the OneKey data.
4. Links based on the percentage of their physicians linked to systems reported in the OneKey data (referred to as the dominant system percentage).

Because the two files use different data sources and methodologies to identify links to health systems, an outpatient site could link to a different health system than the physician group practice that delivers care at that site.

E. Limitations of RUCA Data

The most recent RUCA codes available as of February 2024 are based on data from the 2010 census and the 2006–2010 American Community Survey. About 25 percent of outpatient sites in our data have census tracts that were newly added in 2020, suggesting that population density in those tracts increased substantially between the 2010 and 2020 censuses.^{xxiii} Therefore, the RUCA codes in these data might not accurately reflect what the RUCA code would be based on 2020 census data.

^{xxii} Details on the group practice linkage file are available at <https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/2018-chsp-tin-linkage-file-tech-doc.pdf>.

^{xxiii} See <https://www2.census.gov/geo/pdfs/reference/GARM/Ch10GARM.pdf> for more information about census tract splits.

Appendix A. Data Dictionary – Outpatient Site Linkage File

Variable Name	Variable Type	Description
compendium_os_id	Character	Unique compendium outpatient site ID
onekey_id	Numeric	IQVIA outpatient site ID from OneKey
os_name	Character	Outpatient site name
os_street1	Character	Outpatient site street address line 1
os_street2	Character	Outpatient site street address line 2
os_city	Character	Outpatient site city
os_state	Character	Outpatient site state
os_zip	Character	Outpatient site street five-digit ZIP
os_zip_plus4	Character	Outpatient site street extended ZIP Code (+4)
os_type	Character	Outpatient site facility type
os_specialty	Character	Outpatient site facility specialty
os_mds	Numeric	Physician count
health_sys_id	Character	Unique Compendium health system ID (assigned by the CHSP Initiative) of the system linked to the hospital (same variable as in the Compendium of U.S. Health Systems)
health_sys_name	Character	Health system name (same variable as in the Compendium of U.S. Health Systems)
health_sys_city	Character	Health system city (same variable as in the Compendium of U.S. Health Systems)
health_sys_state	Character	Health system state (same variable as in the Compendium of U.S. Health Systems)
corp_parent_name	Character	Outpatient site corporate parent owner ID from OneKey
corp_parent_id	Numeric	Outpatient site corporate parent owner name from OneKey
corp_parent_type	Character	Outpatient site corporate parent owner type from OneKey
tract	Character	Federal Information Processing Standards state + county + census tract code
county_sub	Character	Federal Information Processing Standards state + county + county subdivision code
county	Character	Outpatient site county name
cbsa_code	Character	Core Based Statistical Area code
cbsa_name	Character	Core Based Statistical Area name
rural_nonmsa	Numeric	Indicator for whether this site is outside of a metropolitan statistical area
ruca_prim	Numeric	Primary rural-urban commuting area code designated by the U.S. Department of Agriculture's Economic Research Service
ruca_sec	Numeric	Secondary rural-urban commuting area code designated by the U.S. Department of Agriculture's Economic Research Service
mua	Numeric	Indicator for whether the outpatient site is in a medically underserved area
mua_id	Character	Unique identifier(s) for the medically underserved area(s) where the outpatient site is located
mup	Numeric	Indicator for whether the outpatient site is in a medically underserved population

Variable Name	Variable Type	Description
mup_id	Character	Unique identifier(s) for the medically underserved population(s) where the outpatient site is located
pchpsa	Numeric	Indicator for whether the outpatient site is in a primary care health professional shortage area
pchpsa_id	Character	Unique identifier(s) for the primary care health professional shortage area(s) where the outpatient site is located

Appendix B. Rural-Urban Commuting Area Code Descriptions

Primary RUCA ^a	Secondary RUCA ^b	Description ^c
1		Metropolitan area core: primary flow^d within an urbanized area (UA)^e
	1	No additional code
	1.1	Secondary flow 30% to 50% to a larger UA
2		Metropolitan area high commuting: primary flow 30% or more to a UA
	2	No additional code
	2.1	Secondary flow 30% to 50% to a larger UA
3		Metropolitan area low commuting: primary flow 10% to 30% to a UA
	3	No additional code
4		Micropolitan area core: primary flow within an urban cluster (UC) of 10,000 to 49,999 (large UC)^f
	4	No additional code
	4.1	Secondary flow 30% to 50% to a UA
5		Micropolitan high commuting: primary flow 30% or more to a large UC
	5	No additional code
	5.1	Secondary flow 30% to 50% to a UA
6		Micropolitan low commuting: primary flow 10% to 30% to a large UC
	6	No additional code
7		Small town core: primary flow within a UC of 2,500 to 9,999 (small UC)
	7	No additional code
	7.1	Secondary flow 30% to 50% to a UA
	7.2	Secondary flow 30% to 50% to a large UC
8		Small town high commuting: primary flow 30% or more to a small UC
	8	No additional code
	8.1	Secondary flow 30% to 50% to a UA
	8.2	Secondary flow 30% to 50% to a large UC
9		Small town low commuting: primary flow 10% to 30% to a small UC
	9	No additional code
10		Rural areas: primary flow to a tract outside a UA or UC
	10	No additional code
	10.1	Secondary flow 30% to 50% to a UA
	10.2	Secondary flow 30% to 50% to a large UC
	10.3	Secondary flow 30% to 50% to a small UC
99		Not coded: Census tract has zero population and no rural-urban identifier information

RUCA = Rural-Urban Commuting Area.

^a The primary RUCA code classifies area rurality based on population density, urbanization, and largest daily commuting patterns.

^b The secondary RUCA code further subdivides primary RUCA codes based on secondary commuting patterns.

^c We obtained the RUCA code definitions from <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/>.

^d Primary flow refers to the largest commuting pattern in the area.

^e A UA as used in this context is an area with a population of 50,000 or more.

^f A UC as used in this context is an area with a population of 2,500 to 50,000.