

# Comparative Health System Performance Initiative: Compendium of U.S. Health Systems, 2016, Hospital 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](http://www.ahrq.gov)

**Contract Number:** HHSA290201600001C

## Prepared by:

Mathematica Policy Research, Washington, DC  
Kirsten Barrett  
Ken Peckham  
David Jones  
Rachel Machta  
Jessica Heeringa  
Eugene Rich

**AHRQ Publication No. 18-0063**  
**October 2018**



**This page intentionally left blank.**

# Contents

- Acknowledgments..... iv
- I. Introduction..... 1
- II. Data Sources ..... 2
- III. Methodology..... 3
  - A. Hospital-to-Hospital Matching ..... 3
  - B. Linking Hospitals to Health Systems..... 5
- IV. Hospital Linkage File Contents ..... 6
- V. Caveats and Limitations..... 7
  - A. Working Definition of Health System ..... 7
  - B. Differences Across Data Sources..... 8
  - C. Multiple Hospitals Sharing the Same CCN ..... 8
  - D. Mergers and Acquisitions ..... 9
- Appendix A. Non-Federal General Acute Care Flag..... 10
- Appendix B. Data Dictionary – Hospital Linkage File..... 15

# Tables

- Table II.1. Potential data sources for health system identification ..... 2
- Table III.1. Match scores and descriptions ..... 4
- Table IV.1. Hospitals in the Compendium hospital linkage file ..... 7
- Table A.1. Definitions of non-Federal general acute care hospitals ..... 10
- Table A.2. Percentage agreement for the non-Federal general acute care hospital flag..... 11
- Table A.3. Comparison of Compendium and AHA general acute care hospital flag..... 11
- Table A.4. Differences in the general acute care hospital flags between the Compendium and AHA, by AHA hospital type..... 12
- Table A.5. AHA hospital types (non-general acute care) among general acute care Compendium hospitals with >50 percent match rate on general acute care hospital flag across data sources..... 12
- Table A.6. Percentage agreement on general acute care hospital flags across data sources (hospitals not in Compendium systems)..... 14

## 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, 2016, Hospital Linkage File:

- Agency for Healthcare Research and Quality (AHRQ) (Mike Furukawa, Dan Miller, Linda Bergofsky, and Jing Guo)
- Comparative Health System Performance (CHSP) Initiative Centers of Excellence (Dartmouth College, National Bureau of Economic Research [NBER], RAND Corporation)
- CHSP Technical Expert Panel (TEP)
- Data vendors (QuintilesIMS™ [now IQVIA™] and the American Hospital Association)
- Additional Mathematica Policy Research staff (Linda Molinari, Jim Reschovsky, and Jessica Coldren)

# I. Introduction

In 2015, the Agency for Healthcare Research and Quality (AHRQ) created the Comparative Health System Performance (CHSP) Initiative to study how health care systems promote evidence-based practices in delivering care.<sup>i</sup> 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. To this end, the initiative established three Centers of Excellence (CoE) and a Coordinating Center to identify, classify, track, and compare health systems. AHRQ established CoEs at Dartmouth College, the National Bureau of Economic Research (NBER), and RAND Corporation.

Mathematica Policy Research serves as the initiative's Coordinating Center. Mathematica is working collaboratively with AHRQ and the CoEs to facilitate synthesis of findings on comparative health system performance, build a Compendium of health system resources, and support the dissemination of the CHSP Initiative findings.

A key first step of the initiative was to identify and enumerate health systems. To achieve this objective, AHRQ developed a Compendium of U.S. Health Systems, which was released in September 2017. The initial release of the Compendium comprised a consolidated list of 626 health systems in the United States in 2016 (referred to as the "list" in this document).

AHRQ developed the list using information on systems from several data sources that identify systems and their members. In addition to the names and locations of the systems, the list includes a set of their attributes (for example, the number of physicians and hospitals in the system and the number of hospital discharges from system hospitals).

In September 2018, AHRQ added a file linking the 626 health systems with their member hospitals (referred to as the "hospital linkage file" in this document). The hospital linkage file includes information on 6,762 hospitals, including hospital name, street address, city, State, and ZIP Code. In addition, the file includes hospital identifiers such as the Centers for Medicare & Medicaid Services (CMS) certification number (CCN) and American Hospital Association (AHA) ID, health system name, and Compendium health system ID. The hospital linkage file is publicly available so users can identify hospitals within health systems and then, using additional data sources, examine aspects of systems and their members, such as cost and quality of care.

The list of health systems and the hospital linkage file are resources for researchers, policymakers, and other stakeholders who want to understand how health systems can improve the value of health care. The Compendium of U.S. Health Systems, 2016, the hospital linkage file, and corresponding technical documentation form the basis of the Compendium resources. These reside on AHRQ's website and will be updated during the course of the initiative with additional resources to support research on health systems.<sup>ii</sup>

---

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

<sup>ii</sup> The Compendium of U.S. Health Systems, 2016, and technical documentation can be found at <https://www.ahrq.gov/chsp/Compendium/index.html>.

This document summarizes the approach taken to create the hospital 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 it. In section IV, we describe the variables contained in the linkage file. The document concludes with a brief list of caveats that should be considered when using the hospital linkage file.

## II. Data Sources

Although many data sources are useful for studying health system components and their attributes, only a handful explicitly identify systems and indicate relationships among organizations that can be used to identify which collections of organizations constitute a system. These data sources include the following:

- Healthcare Organization Services (HCOS) database<sup>iii</sup>
- SK&A Healthcare Databases
- AHA annual survey of hospitals

We use these three data sources because they identify hospitals, link hospitals to systems, and are nationally representative. Table II.1 briefly describes each data source, including the name of the data holder, intended purpose of the data, and health system and hospital identifiers.

**Table II.1. Potential data sources for health system identification**

Source	Data Holder	Intended Purpose	System Identifiers	Hospital Identifiers and Hospital Count
Healthcare Organization Services (HCOS), December 2016	QuintilesIMS (now IQVIA)	Reference database for sales/marketing purposes	Integrated delivery network	CCN n=5,732
SK&A Healthcare Databases, December 2016	QuintilesIMS/ SK&A (now IQVIA)	Reference database for sales/marketing purposes	Integrated health system	AHA ID n=6,992
America Hospital Association (AHA) Annual Survey, 2015	AHA	Membership database for policy research and industry monitoring	Hospital system	AHA ID & CCN n=6,251

HCOS and SK&A maintain frequently updated databases that contain information on health systems, physicians, and hospitals nationwide. HCOS contains system- and facility-level data on staffing, beds, and facility type. Data are collected through a combination of telephone surveys and administrative sources. The data describe relationships that providers have with hospitals and group practices via ownership, management, leasing, purchasing, and contracting mechanisms. SK&A contains provider-, facility-, and system-level data on professional training

---

<sup>iii</sup> IQVIA maintains two integrated databases relevant to the study of health system performance under the umbrella of Healthcare Relational Services: OneKey Organizations, formerly known as HCOS, and OneKey Professionals, formerly known as HCPS. Throughout the document, we refer to OneKey Organizations as HCOS.

and specialty, location, ownership, and affiliations. AHA data are based on an annual census of hospitals. An overview of the data collection methods for HCOS, SK&A, and AHA can be found in the Compendium of U.S. Health Systems, 2016, Technical Documentation.

### III. Methodology

The AHA, HCOS, and SK&A data sources contained 6,251, 5,732, and 6,992 hospitals, respectively. In this section, we first describe the hospital-to-hospital matching approach to link hospitals across the three data sources to create a single set of unique hospitals. We then describe the approach to link hospitals to systems.

#### A. Hospital-to-Hospital Matching

To identify a set of unique hospitals across the three data sources, we conducted a series of six paired matches:

1. AHA-to-HCOS,
2. AHA-to-SK&A,
3. SK&A-to-HCOS,
4. SK&A-to-AHA,
5. HCOS-to-AHA, and
6. HCOS-to-SK&A.

Within each paired match, we prioritized matches based on CCN or AHA ID. We also identified matches based on hospital names and locations, which we considered lower quality matches than those based on hospital IDs. We describe each step in more detail below.

**Step 1.** We first matched AHA-to-HCOS and AHA-to-SK&A data because the AHA database is the only source of the three that contains both the CCN and AHA ID. Thus, it not only matches directly to the HCOS and SK&A data, but also forms a bridge between these two data sources. We identified AHA-to-HCOS matches based on CCN and AHA-to-SK&A matches based on AHA ID.<sup>iv</sup>

We also used the SAS COMPGED function to determine the extent to which hospital names matched. COMPGED assigns a score to each potential match, which reflects the degree to which hospital names match alphanumerically. A COMPGED score of 0 reflects an exact name match; a higher score indicates a lower quality match. Finally, we examined the geographic proximity of hospitals between the paired data sources based on geocoded street address, city, State, and ZIP Code; and exact text matches on city, State, and ZIP Code.

The matching often resulted in a one-to-many match for each pairwise comparison (except for CCN and AHA ID matches); for example, many AHA hospitals had multiple potential HCOS hospital matches based on name or geography. We assigned a match score to each potential

---

<sup>iv</sup> Before matching hospitals, we deduplicated SK&A by AHA ID and HCOS by CCN, choosing the record closest to the matched AHA hospital based on AHA ID or CCN. This approach resulted in the removal of 99 SK&A and 169 HCOS hospital records.

match in order of the quality of the match (lower scores indicate higher quality matches), as shown in Table III.1.

All hospital pairs with a match score of 1 (CCN or AHA ID matches) were assessed as “matched” and set aside as matches. We also set aside hospital pairs with match scores of 2 or 3 as matches. If a hospital was matched to multiple hospitals with scores between 1 and 3, we chose the match with the lowest score. All other hospitals remained in the pool of unmatched hospitals as possible matches in the next steps.

**Table III.1. Match scores and descriptions**

Match Score	Description
1	Exact match on CCN or AHA ID
2	COMPGED $\leq$ 150 and (ZIP Codes match or street names match and hospitals are within 1 mile of each other)*
3	Hospitals are within one-half mile of each other
4	COMPGED $\leq$ 150 on full name (e.g., Jones Medical Center)
5	COMPGED $\leq$ 150 on truncated hospital name (e.g., Jones Medical)
6	Hospitals are within 10 miles of each other
7	Exact numeric match on ZIP code
8	Exact text match on city and State

\*HCOS was the only data source among the three that had both a mailing address and physical address. For purposes of matching, we used the HCOS physical address.

**Step 2.** We repeated the same process of matching each hospital in the SK&A data file to AHA and HCOS hospitals. As in Step 1, this step resulted in a number of one-to-many matches, and we assigned a match score to each potential match. Scores ranged from 2 to 8 (because all CCN and AHA ID matches were removed earlier). We sorted matched hospitals by match code and used the lowest match code to determine if the match would be considered a match (match score of 2 or 3) and removed those hospitals from subsequent pairwise comparisons.<sup>v</sup>

**Step 3.** We repeated the preceding step, matching each hospital in the HCOS data file to AHA and SK&A hospitals. As in Steps 1 and 2, this step resulted in a number of one-to-many matches. We assigned a match score to each potential match and sorted them. Match scores of 2 and 3 were considered a match.<sup>vi</sup>

---

<sup>v</sup> Based on a manual review of the quality of matches by match score, we determined that we could be reasonably confident in match scores 2 and 3, but there was a noticeable decline in our confidence in the matches beginning with match scores above 3.

<sup>vi</sup> Ninety-two hospitals that matched on name or address had conflicting information for one of the hospital IDs. For example, a hospital in the AHA data might match a hospital in the HCOS data based on CCN and a hospital in the SK&A data based on hospital name, but the AHA IDs in the AHA data and SK&A data differ. Of these hospitals, 55 had matching AHA IDs but different CCNs, and 37 had matching CCNs but different AHA IDs. An additional 14 hospitals matched on name or address but did not match on CCN or AHA ID. In these cases, we report the CCNs from the AHA data.



We identified 7,786 unique hospitals among the three data sources. We excluded 1,024 hospitals from the hospital linkage file that did not report a CCN or an AHA ID. We excluded them because, although they did not match any other hospitals based on name or address, without a CCN or AHA ID, we cannot be confident that they are unique hospitals.

In addition, without a CCN or AHA ID, it will be difficult for users of the linkage file to confidently match these hospitals to other data sources for analysis. This exclusion resulted in a hospital linkage file with 6,762 unique hospitals. Among those hospitals, 5,981 matched across at least two of the three data sources on the CCN or AHA ID. An additional 343 hospital matches were the result of matching on name and geographic proximity (match code 2 or 3). Finally, 438 hospitals in the hospital linkage file did not match with a hospital from another source, based on CCN, AHA ID, name, or location.

## **B. Linking Hospitals to Health Systems**

As described in the Compendium of U.S. Health Systems, 2016, Technical Documentation, hospitals need to be assigned to health systems for two reasons.<sup>vii</sup> First, health systems contained in the Compendium are required to have at least one non-Federal general acute care hospital. Second, the list of health systems includes system attributes that are hospital-level variables aggregated to the system level, such as:

- Number of hospitals per system,
- Number of general acute care hospitals per system,
- Extent to which system hospitals are located in multiple States, and
- Number of hospital discharges per system.

Each of the three data sources identifies hospitals included in systems. Before linking hospitals to health systems, we excluded those that did not report a CCN or an AHA ID. To assign the remaining 6,762 hospitals to health systems, we applied the following rules:

1. We linked hospitals reported as being assigned to only one health system to that system.<sup>viii</sup> Of the 6,762 unique hospitals with a CCN or AHA ID reported across the three data sources, 5,510 hospitals were assigned to at least one health system.<sup>ix</sup> Of the 5,510 hospitals, 5,112 were assigned to one health system only, and 398 were assigned to two or more systems. The remaining 1,252 hospitals (18.5 percent) were not assigned to any health system.
2. We manually reviewed the 398 hospitals identified as being assigned to two or more health systems due to discrepancies across data sources. In most cases, the multiple

---

<sup>vii</sup> Available at <https://www.ahrq.gov/chsp/Compendium/technical-documentation.html>.

<sup>viii</sup> Hospitals linked to one health system include those in which all three data sources agree on the system assignment, as well as cases in which one or more of the data sources agree and one or more do not list the hospital in a system. For example, if one of the data sources assigns a hospital to a system, and the other two data sources do not assign a system for the hospital, it is listed as being in the system. For 4,574 of the 5,510 hospitals in systems, all three sources agreed on the system assignment; for 634 hospitals, two sources agreed on the system assignment; and for 302, only one source listed the hospital as being assigned to the system.

<sup>ix</sup> Note, this number of hospitals is prior to excluding systems that do not meet the Compendium definition of a health system, which is discussed later in this section.

systems were in fact the same system with a different name or systems nested within each other (that is, subsystems and parent systems). In the former cases, we updated the list to indicate that these systems were a match and the hospital belonged to this system; in the latter case, we assigned the hospital to the parent system. We also identified cases in which a change in ownership occurred, which was reflected in one data source but not another. We updated these linkages to reflect the change.

3. The remaining cases were hospital-level joint ventures, in which multiple systems have a formal relationship with a hospital. In these cases, we linked the hospitals to a system using the following decision rules, in order:
  - When it was clear, based on a manual review of systems' and hospitals' websites, that one system was the majority owner or taking responsibility for running the day-to-day operations of the hospital, we linked the hospital to that system.
  - We linked the hospital to the system based on investor-owned status; that is, if the hospital was investor owned as reported in the Healthcare Cost Report Information System data, and only one of the systems was investor owned, we linked the hospital to that system.
  - In the absence of other information, we linked the hospital to the system whose headquarters location was closest to the location of the hospital; that is, we aimed to link it to a local system over regional or national systems.

After applying the exclusion criteria to remove systems that do not meet the Compendium definition of a health system and linking hospitals in subsystems to their parent systems (for a detailed description of this process, see page 17 of the Compendium of U.S. Health Systems, 2016, Technical Documentation), we had 3,949 hospitals linked to the 626 systems in the Compendium of U.S. Health Systems, 2016.

The result of linking hospitals to systems is a hospital linkage file that contains one row for each hospital with a CCN or AHA ID, along with its identifying information (CCN, name, and address). In addition, if the hospital is in a Compendium system, the file includes the identifying information of its system. No hospital is linked to more than one health system.

## **IV. Hospital Linkage File Contents**

The hospital linkage file contains 6,762 hospitals (Table IV.1). A total of 3,949 hospitals are linked to Compendium health systems, of which 3,513 are non-Federal general acute care hospitals.<sup>x</sup> The hospital counts within the Compendium list of health systems sum to match the counts linked to systems in the hospital linkage file. An additional 2,813 hospitals contained within the hospital linkage file are not linked to a Compendium health system; 1,528 of these are non-Federal general acute care hospitals. The remaining 1,285 hospitals are specialty hospitals such as surgical, long-term acute care, and rehabilitation hospitals.

---

<sup>x</sup> See Appendix A for a description of how we identified non-Federal general acute care hospitals.

**Table IV.1. Hospitals in the Compendium hospital linkage file**

<b>Hospital Group</b>	<b>N</b>
<b>Total hospitals in Compendium linkage file</b>	<b>6,762</b>
<b>Linked to Compendium health systems</b>	<b>3,949</b>
Non-Federal general acute care hospitals	3,513
<b>Not linked to Compendium health systems</b>	<b>2,813</b>
Non-Federal general acute care hospitals	1,528

The hospital linkage file contains 12 variables, including each hospital’s name; location (address, city, State, and ZIP Code); a flag for non-Federal general acute care hospitals; a unique hospital ID assigned as part of the development of the Compendium; and, if applicable, their linked Compendium health system ID, name, and location.<sup>xi</sup>

The hospital linkage file contains two linking variables. Medicare CCN (ccn) can be used to link hospitals to external data sources. The unique health system identifier (health\_sys\_id) can be used to link the hospital to the systems in the Compendium of U.S. Health Systems. The linkage file does not include the AHA ID, as release of this variable is not permitted under our AHA data use agreement. Appendix B contains a data dictionary for the hospital linkage file.

## **V. Caveats and Limitations**

This initial release of the hospital linkage file enables users to link Compendium health systems with their member hospitals. When using the list, users should bear in mind a few caveats and limitations to the current methods for assigning hospitals to health systems. AHRQ will continue to work closely with the CoEs, the Coordinating Center, the CHSP Technical Expert Panel, and other stakeholders to identify strategies to address these limitations in future versions of the hospital linkage file.

### **A. Working Definition of 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.<sup>xii</sup>*

Hospitals within health systems that do not meet this working definition are not identified in the hospital linkage file as being part of a system. Also, hospitals in the hospital linkage file are

---

<sup>xi</sup> When hospital names or locations differed across data sources, we reported information in the hospital linkage file from AHA, SK&A, and HCOS, in priority order.

<sup>xii</sup> 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.

associated with parent systems; relationships between hospitals and subsystems are not identified.

## **B. Differences Across Data Sources**

The three data sources used in the development of the hospital linkage file vary in ways that affect their characterization of hospitals and their linkages to systems. The sources use different data collection methods; also, AHA data were from 2015, whereas those from SK&A and HCOS were from 2016.

We attempted to maximize the coverage of hospitals in the United States by using multiple data sources. However, some hospitals may be missing from the three data sources, and thus would not be reflected in the hospital linkage file. Similarly, we used AHA, SK&A, and HCOS to determine whether hospitals were designated as a non-Federal general acute care hospital. However, differences in this designation across data sources may result in some general acute care hospitals not being flagged as such or vice versa (see Appendix A).

Finally, the mechanism and criteria for assigning hospitals to health systems differ by data source, which leads to differences in linkage. We encountered differences in health system assignments across data sources that we had to adjudicate. Also, we take a fairly inclusive approach, only requiring that one data source indicate a hospital is assigned to a system.

## **C. Multiple Hospitals Sharing the Same CCN**

The Compendium hospital linkage file does not include hospitals reporting the same CCN (for example, separate hospital facilities or campuses with different sites of care under the same parent hospital) as separate entries in the hospital linkage file; rather, the file includes a single entry for the CCN in these cases. Reporting by CCN, and thus reporting some hospitals together in a single entry in the linkage file, is a limitation for those aiming to identify all individual hospital facilities and campuses and their linkages to health systems. However, there are data sources available (for example, the AHA units file) that split some CCNs into multiple hospitals. Users of the linkage file that have access to data that disaggregates hospitals within CCNs can use the linkage file to assign the multiple hospitals under a CCN to the correct system; for example, if CCN 111 includes hospital 222 and 333, and the CCN is in a system, hospitals 222 and 333 can be assigned to the system. This approach assumes that all hospitals under a CCN are linked to the same system.<sup>xiii</sup>

---

<sup>xiii</sup> We include hospitals with AHA IDs but missing CCNs, many of which are those found in the SK&A data; upon linking them to the AHA units file, we found that many were hospitals that have a CCN but were split into multiple AHA IDs. There are 548 hospitals without CCNs but with AHA IDs in the hospital linkage file; 288 are linked to Compendium health systems, and of these, 179 are non-Federal general acute care hospitals. Approximately two-thirds of the 548 hospitals were found only in the SK&A data (n=360); among these 360, 235 were found in the AHA units file. The AHA units file breaks some CCNs and AHA IDs into separate hospitals, meaning that the hospital is part of a larger parent AHA ID that also has a CCN. The remaining 188 hospitals without CCNs are contained in the AHA database, but they lack a CCN; 157 of these hospitals (84 percent) are Federal or non-general acute care hospitals. Many of the 157 (although not all) are Federal hospitals or children's specialty hospitals. It is possible these facilities do not have a CCN because they do not bill Medicare.

## **D. Mergers and Acquisitions**

The list reflects hospitals in the United States at the end of 2016. However, because HCOS, SK&A, and AHA data vary in the periodicity of their updates, there may be lags in updating changes to systems and their hospitals, such as mergers, acquisitions, and name changes, that occurred before the end of 2016. Thus, the period represented by the hospital linkage file aligns with the periods covered by the data sources (2015 for AHA and 2016 for HCOS and SK&A), with some level of updating accomplished through manual review of hospital-system linkages that varied by data source.

## Appendix A. Non-Federal General Acute Care Flag

In the Compendium of U.S. Health Systems, 2016, all health systems were required to have at least one non-Federal general acute care hospital. The hospital linkage file includes a flag that identifies non-Federal general acute care hospitals—those in Compendium systems and single hospitals that are not part of a system. In this appendix, we provide additional details about the construction of the non-Federal general acute care hospital flag and present analyses of how the designation varied across data sources.

### A. General Acute Care Hospital Definition

The non-Federal general acute care hospital flag in the hospital linkage file denotes hospitals that were identified as non-Federal general acute care hospitals in one or more of the three data sources (AHA, HCOS, and SK&A) used to develop the Compendium of U.S. Health Systems, 2016. Table A.1 describes how non-Federal general acute care hospitals were identified in each data source.

**Table A.1. Definitions of non-Federal general acute care hospitals**

Data Source	Description of Non-Federal General Acute Care Hospitals
AHA	(1) Not Air Force, Army, Navy, Public Health Service, Veterans Affairs, Federal other, Indian Health Service, or Department of Justice facilities and (2) the hospital provides one or more of the following services: general medical and surgical or children's general medical and surgical care or is a critical access hospital or a major or minor teaching hospital based on resident to bed ratio.
HCOS	(1) Not a government or Veterans Affairs-owned/run facility and (2) flagged as an acute care general hospital, critical access hospital, or children's hospital.
SK&A	(1) Not military or Veterans Affairs-owned/run facility and (2) flagged as a general hospital, critical access hospital, or children's hospital.

If a hospital was considered a general acute care hospital in at least one of the three data sources, we flagged it as such in the hospital linkage file.

### B. General Acute Care Hospitals Within Compendium Systems

We examined the differences in the non-Federal general acute care flags across the three data sources and considered alternative approaches to defining the overarching flag reported in the hospital linkage file. Using the approach described above, we identified 3,513 non-Federal general acute care hospitals in the AHA, HCOS, and SK&A data that are linked to Compendium health systems. We also added a non-Federal general acute care hospital flag from a fourth data source—Healthcare Cost Report Information System—to assess its alignment with the three data sources used to create the hospital linkage file.

For each hospital, we calculated the percentage of the four flags indicating it is a general acute care hospital (percentages are calculated only among nonmissing flags; for example, if three of the four flags are nonmissing, and two of the three nonmissing flags indicate a general acute care hospital, the percentage is 66.7 [2/3]). Among hospitals in Compendium health systems flagged as general acute care hospitals (n=3,513), only 25 hospitals have less than 50 percent agreement (Table A.2), which reflects less than 1 percent of all hospitals in Compendium health systems.

Among the remaining 3,488 hospitals, the vast majority (96.1 percent) had 75 percent or greater agreement with regard to their general acute care hospital flags across data sources.

**Table A.2. Percentage agreement for the non-Federal general acute care hospital flag**

Percentage Agreement on General Acute Care Hospital Flag Across Data Sources	Frequency	Percent
25.0	19	0.54
33.3	6	0.17
50.0	97	2.76
66.7	16	0.46
75.0	250	7.12
100.0	3,125	88.96
Total	3,513	100.00

### C. Differences in the General Acute Care Hospital Flags

Because the AHA data include the most detail on hospital types, we examined the extent to which the general acute care hospital flags aligned with that data source. There are 117 hospitals for which the flag in the hospital linkage file indicates that a hospital is a non-Federal general acute care hospital but the AHA flag does not (Table A.3). These are the only 117 hospitals for which the Compendium and AHA flags disagree (apart from the 419 hospitals in the hospital linkage file that do not appear in the AHA data). There were no instances in which the AHA data flagged a hospital as a general acute care hospital while the Compendium did not.

**Table A.3. Comparison of Compendium and AHA general acute care hospital flag**

Compendium General Acute Care Hospital Flag	AHA General Acute Care Hospital Flag			Total
	0	1	Missing	
0	290	0	146	436
1	117	3,123	273	3,513
Total	407	3,123	419	3,949

A further look at the 122 hospitals with 50 percent or less agreement among general acute care hospital flags reported in Table A.2 shows that 94 are listed as non-general acute care hospitals in the AHA data (Table A.4). Of the remaining 28 hospitals, 13 are listed as general acute care hospitals in the AHA data and 15 are not found in the AHA data.

**Table A.4. Differences in the general acute care hospital flags between the Compendium and AHA, by AHA hospital type**

<b>AHA Hospital Type: Non-General Acute Care (N=94)</b>	<b>N</b>	<b>Percent</b>
Surgical	25	20.49
Orthopedic	14	11.48
Children's orthopedic	14	11.48
Heart	11	9.02
Cancer	7	5.74
Rehabilitation	6	4.92
Acute long-term care hospital	4	3.28
Eye, ear, nose, and throat	3	2.46
Children's rehabilitation	3	2.46
Psychiatric	3	2.46
Children's chronic disease	1	0.82
Children's other specialty	1	0.82
Other specialty treatment	1	0.82
Obstetrics and gynecology	1	0.82
<b>AHA Hospital Type: General Acute Care Hospital</b>	<b>N</b>	<b>Percent</b>
General medical and surgical	13	10.66
Not found within AHA data source	15	12.30
Total	122	100.00

**Note:** Compendium hospitals with  $\leq 50$  percent match on general acute care hospital flag.

There are an additional 23 hospitals in which the percentage agreement is above 50, but the AHA flag differs from the current Compendium general acute care hospital flag. The hospitals are identified as different types in the AHA data, with the most common being surgical, obstetrics and gynecology, other specialty treatment, and children's other specialty (Table A.5).

**Table A.5. AHA hospital types (non-general acute care) among general acute care Compendium hospitals with >50 percent match rate on general acute care hospital flag across data sources**

<b>Hospital Type</b>	<b>N</b>	<b>Percent</b>
Surgical	5	21.74
Tuberculosis and other respiratory diseases	1	4.35
Heart	1	4.35
Obstetrics and gynecology	7	30.43
Orthopedic	1	4.35
Other specialty treatment	4	17.39
Children's other specialty	3	13.04
Acute long-term care hospital	1	4.35
Total	23	100.00

## **D. Alternative General Acute Care Hospital Flag**

We examined alternative approaches to defining the general acute care hospital flag based on the extent to which the flags agreed across two or more data sources. If we defined general acute care hospitals as those in which 50 percent or more of the nonmissing flags indicated that the hospital was a general acute care hospital, 25 Compendium hospitals would no longer be



identified as general acute care hospitals (see Table A.2). This change would result in the loss of one Compendium health system because the only hospital would no longer be identified as a general acute care hospital and thus would no longer meet the Compendium definition of a health system.

Alternatively, if we changed the current flag to be consistent with the AHA general acute care hospital flag whenever it was available, 117 Compendium general acute care hospitals would have their general acute care hospital flag removed (Table A.3). This removal would result in the loss of two Compendium health systems because their only hospitals would no longer be identified as general acute care hospitals.

If we used the AHA flag when available, and when it was not available, identified general acute care hospitals as those for which 50 percent or more of the nonmissing flags indicated it was a general acute care hospital, we would change the flag for 132 hospitals. Therefore, we would lose the same two systems.

In summary, there are a small number of hospitals in Compendium systems and in the hospital linkage file for which the data sources do not agree on whether they are non-Federal general acute care hospitals. While the hospital linkage file includes the flag for non-Federal general acute care hospitals, users of the Compendium can link the file to the data sources discussed in this document or others (for example, CMS's Provider of Services file<sup>xiv</sup>) to define their own flag and other hospital characteristics.

## **E. General Acute Care Hospitals Not in Compendium Systems**

Our analysis identified 1,528 non-Federal general acute care hospitals that are not linked with Compendium health systems. Among these hospitals, 59 (3.9 percent) have less than 50 percent agreement across their general acute care hospital flags (Table A.6). Among the remaining 1,469 hospitals, nearly 90 percent had 75 percent or greater agreement with regard to their general acute care hospital flags across data sources.

---

<sup>xiv</sup> Available at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/index.html>.

**Table A.6. Percentage agreement on general acute care hospital flags across data sources (hospitals not in Compendium systems)**

<b>Percentage Agreement on General Acute Care Hospital Flags</b>	<b>Frequency</b>	<b>Percentage</b>
25.0	36	2.4
33.3	23	1.5
50.0	48	3.1
66.7	86	5.6
75.0	131	8.6
100.0	1,204	78.8
Total	1,528	100.0

Among hospitals not in Compendium health systems that are currently flagged as general acute care hospitals, 135 are not flagged as such in the AHA data. Furthermore, 164 additional hospitals (29 general acute care hospitals and 135 non-Federal general acute care hospitals) are not found in the AHA data.

## Appendix B. Data Dictionary – Hospital Linkage File

Variable Name	Variable Type	Description
compendium_hospital_id	Character	Unique hospital ID created by the CHSP Initiative and used across all CHSP analyses
Ccn	Character	CMS certification number (hospital ID)
hospital_name	Character	Hospital name
hospital_street	Character	Hospital street address
hospital_city	Character	Hospital city
hospital_state	Character	Hospital State
hospital_zip	Character	Hospital ZIP Code
acutehosp_flag	Numeric	Flag for non-Federal general acute care hospitals
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)