Comparative Health System Performance Initiative: Compendium of U.S. Health Systems, 2021, Group Practice 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
David Jones
Michael Barna
Sandi Nelson
Andrew McGuirk
Eugene Rich

AHRQ Publication No. 25-0011-5-EF
December 2024

Agency for Healthcare Research and Quality

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

Contents

Tables

Figures

Acknowledgments

We would like to acknowledge a number of organizations and individuals who provided support and guidance during the development of the Compendium:

- Agency for Healthcare Research and Quality (AHRQ): Dan Miller, Leeann Comfort, Jay Crosson, Lingrui Liu, Mike Furukawa, Bruce Seeman, and Doreen Bonnett
- CHSP Initiative Centers of Excellence: Dartmouth, NBER, RAND
- CHSP Technical Expert Panel
- Data vendors: IQVIA™ and American Hospital Association
- Mathematica staff: Miaomiao Shen, Genna Cohen, and Caitlyn Newhard

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

Executive Summary

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. AHRQ's goal is to understand the factors that affect health systems' use of patient-centered outcomes research (PCOR) and to identify best practices in disseminating and using PCOR.

AHRQ supports the ongoing work of the CHSP Initiative by providing this 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 number of publicly available data resources for researchers, policymakers, and other interested parties who want to understand how health systems can improve the value of healthcare.

To date, CHSP data resources include:

- The Compendium of U.S. Health Systems (lists of health systems in the United States, referred to as the systems lists in this document). [i]
- Hospital linkage files that link hospitals to health systems (referred to as the hospital linkage files in this document).
- Group practice linkage files that link group practices to systems.
- Outpatient site linkage files that link outpatient sites (for example, medical group practice sites) to health systems.
- Nursing home and home health care organization linkage files that links nursing homes and home health care organizations to health systems.

These resources reside on AHRQ's website and are being updated over the course of the initiative with additional resources to support research on health systems.

In October 2019, AHRQ first added a 2016 physician group practice linkage file, which links group practices to their 2016 systems based on information in the data sources that indicate ownership or tight management relationships. Similar to the hospital linkage file, the group practice linkage file enables users to link 2016 data on group practices to the 626 health systems in the 2016 systems list and other data sources.

[i] The systems lists identify health systems with 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.

The group practice linkage file includes group practices both in systems and not in systems.[ii] In December 2020, AHRQ added a 2018 physician group practice linkage file, which links group practices to their 2018 systems.[iii]

In December 2024, AHRQ added physician group practice linkage files for 2020, 2021, and 2022, to enable users to link data on group practices to the health systems in each of the respective systems lists and to other sources. This document describes the methodology and approach used to develop the 2021 group practice linkage file.

A. Methodology

To identify the set of physician group practices (and their physicians) to be considered for linkages to the 2021 Compendium of U.S. Health Systems, we used version 2.4 of the 2021 Medicare Data on Provider Practice and Specialty (MD-PPAS). The file contains records for all providers with a valid National Provider Identifier (NPI) that submitted a Medicare fee-for-service Part B noninstitutional claim with a positive allowed amount in 2021.

We used 2021 Medicare Advantage data to create a file of NPIs linked to Tax Identification Numbers (TINs) billing Medicare Advantage but not Medicare Part B. We used this file to supplement the fee-for-service MD-PPAS file discussed above.

We restricted the set of group practices to those with two or more physicians (medical doctors or doctors of osteopathy, identified by their NPIs), which excluded 71,049 single-physician practices.[iv] These physician group practices give us the set of group practices and their physicians to be considered for linkages to systems: 36,202 group practices with 621,717 physicians.

The group practice linkage file links TINs to health systems. TINs are entities that physicians assign their rights to for submitting Medicare claims and collecting payment for services. We follow the Centers for Medicare & Medicaid Services (CMS) in referring to TINs as group practices, and we use these terms interchangeably in this document.

We identified four approaches to link group practices to candidate health systems and then combined the information gleaned from the approaches to assign the final linkages of group practices to systems. The four approaches are:

1. **CMS Certification Number (CCN) approach:** We used shared TIN information in the 2021 Medicare Provider Enrollment, Chain, and Ownership System (PECOS) to link group practices to hospital CCNs and then used the 2021 hospital linkage file to link the group practices to systems. Using the PECOS data, we determined whether the group

[ii] The technical documentation for the 2016 group practice linkage file can be found at https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/CHSP_2016_GPLF_tech_Dec2020.pdf.
[iii] The technical documentation for the 2018 hospital linkage file can be found at https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/2018-hospital-linkage-techdoc-cx.pdf.
[iv] We also excluded TINs without a TIN name in the 2021 MD-PPAS v2.4 file, which excluded 89 TINs with 214 physicians from the linkage file. Because we cannot report the actual TIN in the public version of the linkage file, TIN name is critical information for users to be able to identify the TIN.

practice TIN was shared with a hospital TIN and then identified the CCN associated with that hospital. We assumed that TINs shared by a group practice and hospital indicated common ownership by the same business entity.

Because we know from the hospital linkage file whether CCNs are linked to systems, we linked the CCNs from the PECOS data to the hospital linkage file. This linkage tells us whether the group practices linked to CCNs are in systems, and for the group practices in a system, it tells us the system's name and Compendium ID.

2. **Hospital-based billing approach:** In this approach, we combined billing information in 2021 Medicare claims data on the settings where physicians provided services (using place of service codes) with information in the hospital linkage file to link group practices to systems. If physicians billing under the same group practice reported that most of their hospital-based services occurred in a facility linked to a specific hospital, we concluded that the group practice had a strong affiliation with the hospital. Furthermore, if the hospital was owned by a health system in 2021, we concluded that the group practice strongly affiliated with the hospital shared the hospital's affiliation with the system.

   Studies of physician-hospital integration have used similar approaches to link physicians to hospitals; for example, by identifying a physician as financially integrated with a hospital if they billed a substantial percentage of outpatient services in a hospital outpatient department (HOPD).[1,2]

   We applied two versions of this approach to link group practices to systems:

   - One that used a group practice's 2021 billings for Medicare beneficiaries occurring at the HOPD settings only and
   - One that used such billings at all hospital-based settings, including inpatient facilities, emergency departments, and HOPDs.

   If most of a group practice's 2021 billings at HOPDs occurred at a specific HOPD linked to a system, we linked the practice to that system. However, basing system assignment on billing through HOPDs may not be informative for physician specialties that provide most of their services at HOPDs regardless of whether they share the system affiliation with the hospitals (for example, anesthesiologists and pathologists). Therefore, we only applied the HOPD criteria described above for group practices in which most physicians are *not* hospital-based specialties.

   In the second version of the approach, we linked group practices with most of their physicians in hospital-based specialties (for example, anesthesiologists and pathologists) to systems based on 2021 billings at *all* hospital-based settings (inpatient facilities, emergency departments, and HOPDs). If such a group practice reported that most 2021 billings at hospital-based settings occurred at specific hospitals linked to a system, we linked the practice to that system.

By including the two versions of this approach, we make two novel contributions to the group practice linkage file: (1) it helps avoid incorrect linkages of hospital-based specialty practices based on the HOPD criteria, and (2) it identifies linkages of hospital-based specialty group practices that are part of systems that would be missed if we did not consider linkages based on billing through all hospital-based settings.

3. **Organizational NPI (Org-NPI) approach:** We used linkages in the 2021 PECOS data between group practices and Org-NPIs combined with information in the 2021 OneKey database from IQVIA to help link group practices to candidate systems.[v] The OneKey data include direct linkages between entities referred to as medical groups in the data (many of which have Org-NPIs) to their health system owners.[vi] Because the PECOS data link group practices to Org-NPIs and the OneKey data link medical groups with Org-NPIs to systems, we could merge the two data sources by Org-NPI.

   We then linked group practices to candidate systems through this Org-NPI link. That is, if a group practice linked to a OneKey medical group through their Org-NPI and the OneKey medical group was in a system, we linked the group practice to the system through this approach.

   Because OneKey also links individual physician NPIs to OneKey medical groups and systems, we can calculate the percentage of the group practice's physicians linked to the system through the Org-NPI approach. We required that either most of the group practice's physicians or at least 50 physicians be linked to the candidate system through the approach to ultimately accept the linkage, thus providing greater confidence that the group practice was linked to the correct system.

4. **Dominant system percentage (DSP) approach:** We combined the NPI-level information on individual physician NPI-system assignments in the 2021 OneKey data with the linkages between physicians and group practices in the 2021 MD-PPAS v2.4 file to help link group practices to systems. Because the MD-PPAS file links TINs to individual NPIs and the OneKey data link individual NPIs to closely affiliated systems, we could merge the two data sources by individual NPI.

   This merge gave us the percentage of a group practice's physicians at the TIN level (in both the MD-PPAS file and the OneKey data) linked to a system. We identified the system with the largest percentage of physicians linked to it as the DSP and considered this system a candidate for potential linkage of the given group practice to that system in 2021.

   When calculating the percentage of a TIN's NPIs linked to systems, we excluded physicians not clearly linked to a single system or group practice: (1) physicians linked to

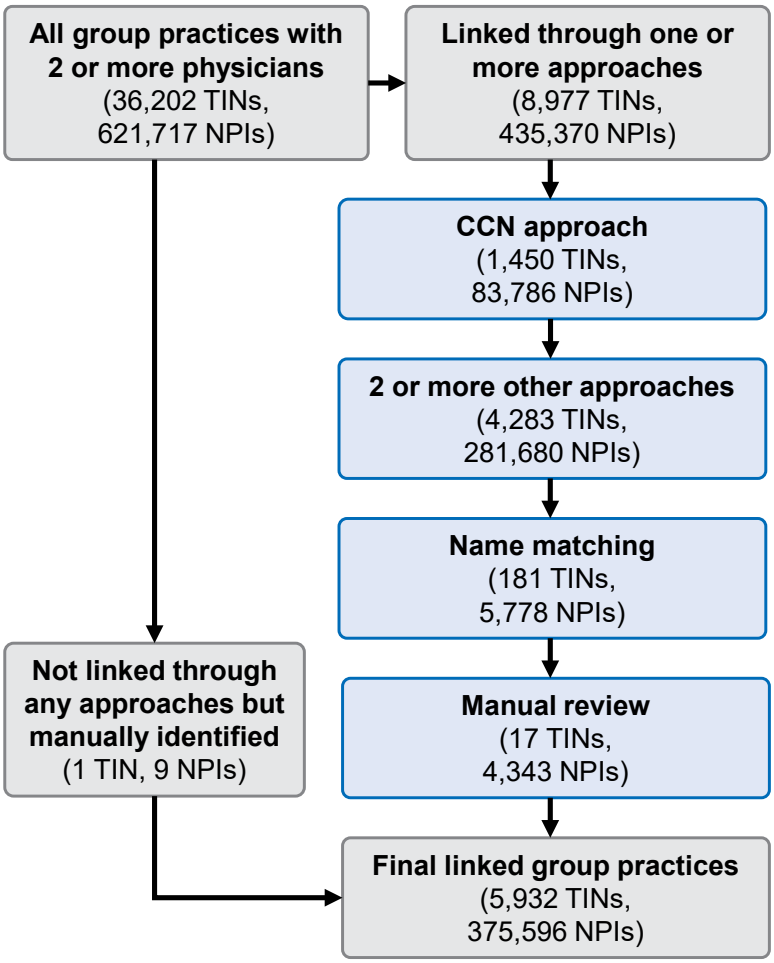
[v] In the 2016 group practice linkage file technical documentation, we refer to the OneKey Organizations data by their former name, the Healthcare Organization Services data.

[vi] OneKey medical groups are typically analogous to practice site locations rather than entire group practices. Thus, while overlap between OneKey medical groups and group practices is useful in linking practices to systems, the two entities are not synonymous.

more than one system in the OneKey data and those linked to a system but also linked to a facility, such as a practice site, not affiliated with a system; and (2) physicians who switched to a different group practice during 2021 in the MD-PPAS v2.4 file. Finally, to provide greater confidence that the group practice was linked to the correct system, we required that either most of the group practice's physicians or at least 50 physicians be linked to the candidate system through the approach.

We used the four types of candidate linkages to make the final group practice linkages in a four-step process (Figure ES.1). Of the 36,202 group practices with 621,717 physicians in the MD-PPAS v2.4 file, 8,977 group practices with 435,370 physicians were linked to a system through one or more of the four approaches.

Figure ES.1. Summary of decision rules for linking group practices to systems



Note: One group practice was not linked using any of the four approaches, but was ultimately linked to a system in the final group practice linkage file through a manual review of systems with no linked group practices.

In the first step, we accepted group practice linkages to a system through the CCN approach, which led to accepted linkages for 1,450 group practices with 83,786 physicians. We accepted all linkages based on the CCN approach, because the linkages are based on only two connections in

which we have relatively high confidence (between the TIN and CCN directly from Medicare enrollment data and the CCN and system that come from the hospital linkage file).

Second, we accepted group practice linkages to a system identified through two or more of the remaining approaches. Linkages made through any one of the approaches could be inaccurate; by requiring corroboration from another approach made through different connections and data, we minimized incorrect linkages. This step led to accepted linkages for an additional 4,283 group practices with 281,680 physicians.

Next, to resolve discrepancies (group practices linked to different systems through two or more approaches), we used group practice-system name matching. We also used name matching to corroborate linkages made through only one of the approaches (excluding the CCN approach). This step led to accepted linkages for an additional 181 group practices with 5,778 physicians.

Finally, we conducted manual review of the following group practices to identify the correct system linkages:

1. 19 large group practices (with 50 or more physicians each—4,834 physicians in total) that still linked to more than one health system after name matching, which led to us accepting linkages for an additional 14 group practices and 4,319 physicians; and
2. Group practices with any associations to five systems that had no group practices linked through the automated portion of the process described above, which led to us accepting linkages for an additional 4 group practices with 33 physicians (representing three of the five systems). One of these 4 practices, with 9 physicians, was not linked using any of the four approaches.

The final group practice linkage file includes linkages for 5,932 of the 36,202 group practices (with two or more physicians) to Compendium systems, which account for 375,596 of the 621,717 physicians in these group practices.

B. Using the Linkage File To Link Group Practices to Other Data Sources

Users of the group practice linkage file can link the data to data sources other than the Compendium through two methods. First, users can link group practice linkage file data to the MD-PPAS data using the following variables common to both (or that can be created using both sources):

- TIN name,
- State where the plurality of the TIN's NPIs are located,
- Total number of physicians in the TIN,
- Total number of nurse practitioners and physician assistants in the TIN, and
- Total number of line items billed through the TIN.

Alternatively, users can connect the data to files derived from PECOS, such as the Public Provider Enrollment data or the Physician Compare National Downloadable File, by using the PECOS Associate Control ID (PAC ID).[vii]

Through any of these linkages, users can identify the physicians in the group practices and thereby the physicians who are tightly affiliated with the Compendium health systems. In turn, users can link information in Medicare claims data to systems by using physician NPIs. By making such linkages possible, the 2021 group practice linkage file enables users to examine the relationships between group practices (and their physicians) and systems and how group practices in systems compare with those not in systems in 2021.

[vii] Refer to https://data.cms.gov/public-provider-enrollment and https://data.medicare.gov/data/physician-compare.

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. AHRQ's goal is to understand the factors that affect health systems' use of patient-centered outcomes research (PCOR) and to identify best practices in disseminating and using PCOR.

AHRQ supports the ongoing work of the CHSP Initiative by providing this 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 number of publicly available data resources for researchers, policymakers, and other interested parties who want to understand how health systems can improve the value of healthcare. To date, CHSP data resources include:

- The Compendium of U.S. Health Systems (lists of health systems in the United States, referred to as the systems lists in this document).[viii]
- System-hospital linkage files that link hospitals to health systems (referred to as the hospital linkage files in this document).
- Group practice linkage files that link group practices to systems.
- Outpatient site linkage files that link outpatient sites (for example, medical group practice sites) to health systems.
- Nursing home and home health care organization linkage files that links nursing homes and home health care organizations to health systems.

These resources reside on AHRQ's website and are being updated over the course of the initiative, with additional resources to support research on health systems.

The Compendium of U.S. Health Systems (referred to as the systems list in this document), first released in 2017, provides a list of health systems in the United States—from 626 systems in 2016 to 635 in 2021—and key system attributes. The systems list consolidates information from several data sources that indicate system ownership and provider affiliations with systems.[ix]

The systems list identifies systems with at least one hospital and at least one group of physicians that collectively provide comprehensive care (that is, including primary and specialty care). The

[viii] The systems lists identify health systems with 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.

[ix] Furukawa, et al.,[3] describe the methodology used to create the systems list and use the systems list to examine health systems in the United States in 2016.

hospitals and physician groups in a system must be connected with each other through common ownership or joint management as identified in the data sources.[x]

The Compendium system-hospital linkage files (referred to as the hospital linkage files in this document) enable users to link health systems with their member hospitals in their respective years.[xi] The hospital linkage files include name and address information on hospitals in systems and not in systems. In addition, the files include hospital and system identifiers such as the Centers for Medicare & Medicaid Services (CMS) certification number (CCN) and Compendium health system ID.

In October 2019, AHRQ added a 2016 group practice linkage file, which links group practices to their 2016 Compendium health systems based on information in the data sources that indicate ownership or tight management relationships.[xii] The group practice linkage file includes information on 41,153 group practices (made up of 556,858 physicians) in systems and not in systems in 2016. Similar to the hospital linkage file, the group practice linkage file contains information to allow users to link the data on the 626 health systems in the systems list to other data sources.

In December 2020, AHRQ added a 2018 physician group practice linkage file, which links group practices to their 2018 systems. The group practice linkage file includes information on 39,103 group practices (made up of 585,012 physicians) in systems and not in systems in 2018.

In December 2024, AHRQ added physician group practice linkage files for 2020, 2021, and 2022, linking group practices to their systems in the respective year. The 2021 file includes information on 36,202 group practices (made up of 621,717 physicians) in systems and not in systems in 2021.

This document describes the methodology and approach used to develop the 2021 group practice linkage file. Section II describes the data sources used. Section III presents the methodology used to create the linkage file. Section IV describes the linkage file contents, and section V presents caveats and limitations.

---

[x] The technical documentation for the 2021 systems list contains more information about how we identified qualifying health systems: https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/2021-Compendium-TechDoc-rev.pdf.

[xi] The technical documentation for the 2021 hospital linkage file is at https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/2021-hospital-linkage-techdoc-rev.pdf.

[xii] When we refer to group practices being part of systems or linked to systems throughout this report, we make the assumption of ownership or tight management relationships between the practices and systems based on the information reported in the data sources and the methodology we developed to link group practices to systems.

II. Data Sources

A. Identifying Group Practices and Physicians Using MD-PPAS

We used version 2.4 of the 2021 Medicare Data on Provider Practice and Specialty (MD-PPAS) data to:

1. Identify the set of physician group practices to be considered for linkages to the 2021 Compendium of U.S. Health Systems and
2. Link physician providers (identified by their National Provider Identifiers [NPIs]) to group practices (identified by TINs), which we use later to help link the group practices to systems.[xiii]

The 2021 fee-for-service MD-PPAS v2.4 file contains records for all providers with a valid NPI that submitted a Medicare fee-for-service Part B noninstitutional claim with a positive allowed amount in 2021.[xiv] We restrict the set of providers considered in the analyses to medical doctors and doctors of osteopathy in the MD-PPAS file, which are identified by their NPIs. We refer to these providers collectively as physicians in this document.

For each physician, the file indicates the two Tax Identification Numbers (TINs) with the largest number of line items billed by the physician in 2021; that is, the primary and secondary TINs the physician used for billing. The data include the physician's specialty, number of line items billed, allowed charges, and unique patients overall and for the physician's primary and secondary group practices. We assigned each physician to his or her primary TIN when developing the group practice linkage file.

In addition, we used 2021 Medicare Advantage (MA) data to create a file of NPIs linked to TINs that bill MA plans but not Medicare Part B. We supplemented the fee-for-service MD-PPAS v2.4 file with MA data to ensure that we included a set of group practices and physicians participating in Medicare that was as comprehensive as possible.[xv] We added MA-only TINs (and their NPIs) to the final set of group practices if the practices were also found in the Medicare Provider Enrollment, Chain, and Ownership System (PECOS) data (described further in the next section).

[xiii] The group practice linkage file links TINs to health systems. We follow the Centers for Medicare & Medicaid Services' practice of referring to TINs, which are entities that physicians assigned their rights to for submitting Medicare claims and collecting payments for services, as group practices. In general, we use the terms TIN and group practice interchangeably in this document. More specifically, we use TIN when referring to the identifier in the data sources (that is, the TIN).

[xiv] Only claims for evaluation and management visits, procedures, imaging services, or nonlaboratory tests are included. More information on the MD-PPAS data is available at https://www.resdac.org/cms-data/files/md-ppas/data-documentation.

[xv] The primary TIN, which is based on the plurality of line items billed, changed for some physicians when we combined the MA data with the fee-for-service MD-PPAS v2.4 file (that is, a different TIN had the plurality of line items when fee-for-service and MA line items were summed).

We required that the group practices also be in the PECOS data to confirm the TINs were valid, because some of the group practice identifiers (TINs) in the MA data were not valid TINs.[xvi] We also added MA-only NPIs to TINs previously identified in the fee-for-service MD-PPAS v2.4 file. This step added 1,040 TINs and 33,511 NPIs that are only in the MA data to the set of TINs and NPIs to be considered for linkages to systems.

Finally, we restricted the set of group practices to those with two or more physicians, which excluded 71,049 single-physician or solo practices (Table II.1).[xvii] These physician group practice linkages provided in the MD-PPAS file give us the set of group practices and their physicians to be considered for linkages to systems: 36,202 group practices with 621,717 physicians.

Table II.1. Number of group practices (TINs) by number of physicians (NPIs)

TIN Size (Number of NPIs)	Number of TINs	Percentage of TINs	Number of NPIs	Percentage of NPIs
All TINs and NPIs	**107,251**	**100.0%**	**692,766**	**100.0%**
1	71,049	66.2%	71,049	10.3%
2+	36,202	33.8%	621,717	89.7%
2-9	27,865	77.0%	101,296	16.3%
10-99	7,391	20.4%	193,388	31.1%
100-499	787	2.2%	156,087	25.1%
500+	159	0.4%	170,946	27.5%

Key: NPI = National Provider Identifier; TIN = Tax Identification Number.

B. Using Other Data Sources To Develop Linkages

We used several other 2021 data sources combined with the 2021 MD-PPAS v2.4 file to assess potential linkages and ultimately make the final linkages between group practices and health systems in the systems list. Table II.2 provides a brief description of each data source, level of information, and available linkage variables. We discuss each data source in more detail below.

Table II.2. 2021 data sources used to link group practices to health systems

Data Source	Description of Information Used in the Data Source	Level of Information	Linkage Variables
MD-PPAS v2.4	Provider-level dataset that assigns Medicare providers (restricted to medical doctors and doctors of osteopathy for the development of the linkage file) by NPI to group practices (TINs)	NPI	NPI, TIN
Hospital linkage file	File that links hospitals (with CCNs) to Compendium systems	CCN	System ID, CCN

[xvi] Invalid TINs appear in the data when a valid TIN was not reported on MA claims.

[xvii] We excluded TINs with a single physician to ensure the confidentiality of identifiable individuals. Physicians who switched TINs during 2021 were not included in the counts of physicians for TINs. We also excluded TINs without a TIN name in either the MD-PPAS v2.4 file or PECOS, which excluded 89 TINs with 214 physicians from the linkage file. Because we cannot report the actual TIN in the public version of the linkage file, TIN name is critical information for users to be able to identify the TIN.

Data Source	Description of Information Used in the Data Source	Level of Information	Linkage Variables
PECOS	System of individual and organizational providers enrolled in Medicare	NPI, TIN, Org-NPI, CCN	NPI, TIN, Org-NPI, CCN
Medicare claims/ hospital-based billing file	Summary of services provided by TIN's physicians at hospital-based settings linked to systems	TIN-CCN, TIN-system	TIN, CCN
OneKey	Database identifying linkages between NPIs and OneKey medical groups, NPIs and systems, and OneKey medical groups and systems	NPI, OneKey medical groups, systems	NPI, Org-NPI

Key: MD-PPAS = Medicare Data on Provider Practice and Specialty; NPI = National Provider Identifier; TIN = Tax Identification Number; CCN = CMS Certification Number; PECOS = Provider Enrollment, Chain, and Ownership System; Org-NPI = organizational NPI.

1. Hospital linkage file

We used information on hospital-to-system assignments in the 2021 hospital linkage file to help identify linkages between group practices and systems. The hospital linkage file contains a record for each hospital and whether the hospital is part of one of the health systems in the systems list. If the hospital is linked to a health system, the file lists the name and identifier of the health system, which links to the systems list.[xviii]

In general, we used the 2021 hospital linkage file (which has hospital CCNs linked to systems) to link group practices to systems once we linked the group practices to hospital CCNs. We describe the approaches using these data in more detail in Section III, Methodology.

2. Medicare Provider Enrollment, Chain, and Ownership System (PECOS)

We used information in the 2021 PECOS data to link TINs to hospital CCNs. Administered by CMS, PECOS is an online system and data repository for individual and organizational providers to enroll in Medicare and manage their enrollments over time. PECOS is not publicly available to researchers, but AHRQ requested permission from CMS to extract selected information from PECOS in order to develop the Compendium of U.S. Health Systems.

We used the enrollment associations (which indicate relationships between entities) in the 2021 PECOS data to make two types of linkages that ultimately helped us link group practices to systems. First, the PECOS data associate group practice TINs with hospital CCNs. In these cases, the group practice shares the same TIN as the hospital, or some physicians providing patient care at the hospital bill Medicare for services through the hospital TIN.

[xviii] More information on the hospitals included in the linkage file and the methodology used to link the hospitals to systems is in the hospital linkage file technical documentation: https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/2021-hospital-linkage-techdoc-rev.pdf.

Second, PECOS data associate TINs with organizational NPIs (Org-NPIs).[xix] The data include one or more Org-NPIs and a TIN for a given group practice enrollment record. Thus, we could link the TIN to the Org-NPIs through the PECOS enrollment record. The relationship between Org-NPIs and TINs is not quite one to one, but 97.9 percent of Org-NPIs link to a single TIN, and 93.6 percent of TINs link to a single Org-NPI. We describe the approach to using these data to link group practices to systems in more detail in Section III, Methodology.

3. Medicare claims

We used billing information in 2021 Medicare fee-for-service claims to create an analytic extract that summarized the extent to which the physicians in group practices provided services at settings of care owned or tightly managed by systems. The analytic extract is referred to in this document as the hospital-based billing file. It includes the number of beneficiary dates of service (DOSs) for Medicare beneficiaries billed through a group practice that occurred at hospital-based settings linked to health systems.[xx]

The hospital-based billing file includes the services (reported on the claims) occurring at: (1) all hospital-based settings, including hospital outpatient departments (HOPDs), inpatient facilities, and emergency departments, and (2) HOPD settings only. For a TIN, the file contains a record for the TIN-CCN combinations; that is, hospitals connected to the practice through billing at hospital-based settings owned by the system. We used this information to assess whether group practices and physicians billing to them should be linked to health systems.

To create the hospital-based billing file, we restricted the data to a set of claims analogous to those used to create the MD-PPAS data, including only claims:

1. With charges for evaluation and management services, procedures, imaging, or nonlaboratory testing,
2. With a valid individual NPI, and
3. With a positive allowed charge amount.

We aggregated the number of DOSs from the claims at the TIN level to generate counts by setting of care (all hospital-based settings and HOPD setting only) for each TIN. The resulting claims-based file indicates, for a given TIN, the number of DOSs for services billed through the TIN.

The hospital-based billing file reports a TIN's count of DOSs separately by CCN, which is indicated on the claims, along with the individual NPI and TIN. The file is at the TIN level with an array of CCNs and the percentage of total DOSs for the corresponding CCN. For example, if 60 percent of the DOSs for a TIN occurred at a facility linked to hospital 1 and 40 percent occurred at a facility linked to hospital 2, the file includes the CCN for hospital 1 in the first

[xix] In the OneKey data, about 54 percent of medical groups have Org-NPIs.

[xx] We use beneficiary dates of service for Medicare beneficiaries as a measure of services provided by setting of care. A date of service can include more than one service and multiple Medicare claims. When developing the 2016 group practice linkage file, we tested the sensitivity of the group practice-system linkages to using line items from claims instead of dates of service, and the results were similar.

CCN field and the CCN for hospital 2 in the second CCN field, along with their corresponding percentage of total DOSs.

Next, we aggregated the TIN-CCN-level information to the TIN-system level. We merged the TIN-CCN file to the hospital linkage file, which contains CCNs linked to systems. We then aggregated the information for each combination of TIN and system to create a hospital-based billing file that contains the count of DOSs billed by setting of care (all hospital-based settings and the HOPD setting only) linked to systems. Thus, the hospital-based billing file contains information on the amount of services (as measured by DOS) billed through TINs occurring at either all hospital-based settings or the HOPD setting only and linked to health systems.

Finally, for each TIN, we calculated the percentage of DOSs with billings that occurred at hospital-based settings that are linked, based on the hospital linkage file, to a given health system. These values give us a measure of the percentage of services billed through a group practice that occurred at hospital-based settings linked to a given health system.

We used this information to identify the system most closely linked to the group practice through this approach; that is, the system with the highest percentage of the TIN's DOSs billed at hospital-based settings. For each TIN, we calculated the percentage of DOSs and identified the system with the highest percentage for all hospital-based settings and the HOPD setting only.

4. IQVIA OneKey database

Finally, we used information on medical group-to-system and physician-to-system assignments from IQVIA's 2021 OneKey database to help identify linkages between group practices and systems.[xxi] The 2021 OneKey data include direct assignments of individual physician NPIs to medical groups and health systems (referred to in OneKey as integrated delivery networks, or IDNs).

OneKey enumerates office-based and hospital-based physicians with close affiliations with facilities owned or managed by an IDN. Physician-to-system affiliations are defined as attending (for system hospitals) and IDN affiliated. Attending includes physicians whose primary practice location is physically in the hospital. IDN affiliated includes physicians who practice at an outpatient location that is part of an IDN campus and admit to one or more IDN hospitals. The system affiliations exclude physicians with admitting privileges at a hospital who are not designated as attending or IDN affiliated.

In addition, the OneKey data include assignments of organizational entities referred to as medical groups (defined as outpatient healthcare centers that provide general or specialized services to patients) to systems that own or tightly manage the groups. OneKey medical groups typically represent practice site locations rather than the entire physician organization. Also, the OneKey data include Org-NPIs for many of the OneKey medical groups the database identifies. These data help us link these OneKey medical groups to TINs identified in PECOS, and

[xxi] In the 2016 group practice linkage file technical documentation, we refer to the OneKey Organizations data by their former name, the Healthcare Organization Services data.

ultimately, the group practices to systems (described in more detail in the Methodology section).[xxii]

[xxii] IQVIA regularly updates information on physicians, medical groups, and systems and the relationships between these entities using web searches, telephone calls with practices, a variety of data sources, and proprietary matching algorithms. More information on the data is available in the Compendium of U.S. Health Systems, 2018, Technical Documentation (https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/2021-Compendium-TechDoc-rev.pdf) and Cohen, et al.[4]

III. Methodology

We combined information from multiple approaches to link group practices to systems to create the 2021 group practice linkage file. In this section, we first describe the various approaches used to identify candidate linkages between group practices and systems. We then describe the decision rules applied to combine the information gleaned from these approaches to make the final assignments of group practices to systems.

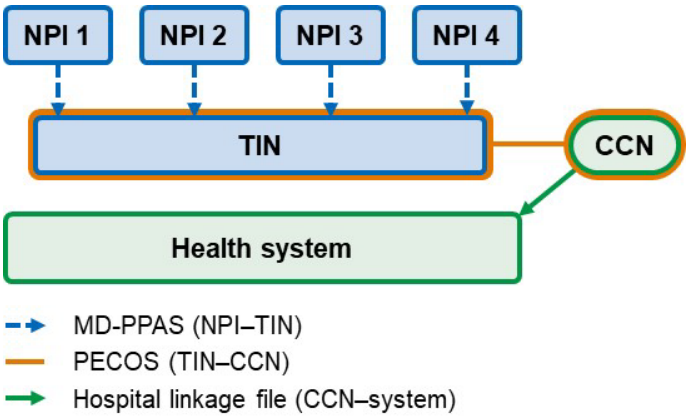
A. Approaches to Linking Group Practices to Systems

1. CMS Certification Number (CCN) approach

This approach used 2021 PECOS data to identify whether the group practice TIN was shared with a hospital TIN. We assumed that TINs shared by a group practice and hospital indicate common ownership by the same business entity. We used the linkages between group practice TINs and hospital TINs (and corresponding CCNs) in the PECOS data and linkages between hospital CCNs and health systems in the 2021 hospital linkage file to link group practices to systems.

In Figure III.1, we show an example of a group practice (TIN) in the MD-PPAS v2.4 file with four physicians (NPIs); the linkages between the NPIs and the TIN are illustrated by the blue dashed arrows. In the PECOS data, some TINs are linked to hospital CCNs (illustrated by the solid orange line linking the TIN to the associated hospital CCN). Next, because we know from the hospital linkage file that the CCN is in a system (illustrated by the solid green arrow), we linked the TIN (and its four physicians—NPIs) to the system through its CCN.

Figure III.1. Illustration of the CCN approach



Finally, we excluded a small number of candidate linkages (4 group practices with 594 physicians), because the CCN approach identified more than one Compendium system linked to the group practice; that is, the TIN links to more than one CCN in PECOS, and the CCNs link to more than one Compendium system.

Of the 36,202 group practices in the MD-PPAS file, 2,368 have a link to CCNs in the PECOS data. Of these group practices sharing a TIN with a hospital, we identified linkages to Compendium systems for 1,450 TINs and their 83,786 NPIs through the CCN approach.

2. Hospital-based billing approach

In this approach, we combined billing information in 2021 Medicare claims data on the settings where physicians provided services using place of service codes (specifically, various types of hospital-based settings) with information in the 2021 hospital linkage file to link group practices to systems. Hospital-based settings included:

- HOPD setting, made up of Off Campus-Outpatient Hospital (place of service code=19) and On Campus-Outpatient Hospital (place of service code=22);
- Inpatient Hospital (place of service code=21); and
- Emergency Room-Hospital (place of service code=23).

We reclassified 2021 Medicare Carrier file claims in the office setting (place of service code=11) as occurring in HOPD settings (place of service code=22) when they:

- Had a matching claim in the Medicare Outpatient file indicating non-emergency room HOPD setting (facility type=1 and type of service=3) and
- Referred to the same patient and service based on matching (1) beneficiary ID, service date, and procedure code or (2) beneficiary ID, service date, and NPI of the service provider.[xxiii]

If physicians billing under the same group practice provided most of their hospital-based services in a facility linked to a specific hospital, it was reasonable to conclude that the group practice had a strong affiliation with the hospital.[xxiv] If the hospital was owned by a health system in 2021, the physician practice strongly affiliated with the hospital may be assumed to share the hospital's affiliation with the system.

We applied two versions of this approach to link group practices to systems: one that used a group practice's services provided at the HOPD setting only and a second that used services at all hospital-based settings. The approach used for a given group practice depends on the specialty mix of the practice. We used services at all hospital facilities for group practices in which most physicians were in hospital-based specialties in 2021. We used services at the HOPD setting for group practices in which most physicians were not in hospital-based specialties in 2021. We discuss the two versions below in more detail.

Linkages based on HOPD billing. This version of the approach relies on the assumption that a group practice billing most of their relevant services in 2021 as having occurred at a specific HOPD linked to a system in 2021 (instead of an office setting owned and managed by their practice) reflects a tight relationship between the practice and the system.[xxv] However, basing

[xxiii] This correction follows the approach used in Neprash, Chernew, and McWilliams.[1]

[xxiv] In this case, relevant services refers to all services provided at hospital-based settings.

[xxv] When a physician service is provided in an ambulatory care setting owned by a hospital (an HOPD), the physician practice receives a greatly reduced payment that covers only the professional component of the service. However, the hospital receives an additional facility fee that often results in the combined payment exceeding what the physician practice would have received for rendering the same service in a physician-owned office setting.[5] When physician practices are purchased by hospitals, services provided by the practice's physicians in the ambulatory care settings formerly owned by the practice can (under some circumstances)[6] be billed to Medicare

system assignment on services at HOPDs may not be informative for physician specialties that provide many or most of their services in an inpatient setting (for example, anesthesiologists and pathologists).

Hospital-based specialties might provide a high percentage of outpatient services at HOPD settings owned by a system regardless of whether the physicians are formally part of the health system. Thus, we used HOPD billing only for group practices in which most of the physicians were not in hospital-based specialties in 2021.

We identified the specialty mix of physicians billing under each group practice in 2021 based on the specialties identified in the 2021 MD-PPAS v2.4 file. We then identified whether most of the TIN's physicians were in specialties that could be classified as hospital based. Hospital-based specialties in the MD-PPAS file include:

- Critical care (intensivists).
- Anesthesiology.
- Nuclear medicine.
- Radiation oncology.
- Emergency medicine.
- Diagnostic radiology.
- Hospitalist.
- Interventional radiology.
- Pathology.

We identified 4,066 group practices as having a majority of physicians in hospital-based specialties (among the 10,600 group practices with some level of linkage to hospital-based settings through billing and services provided).

We accepted group practice linkages to Compendium systems when most of the group practice's DOSs at HOPDs occurred at specific HOPDs linked to a Compendium system and at least 10 DOSs for Medicare beneficiaries were billed at HOPDs in 2021. We applied the final restriction to require that a meaningful number of services be delivered through HOPDs. This approach added to the confidence that the linkages to systems through HOPDs reflected ownership or tight management relationships between the group practice and the linked system.

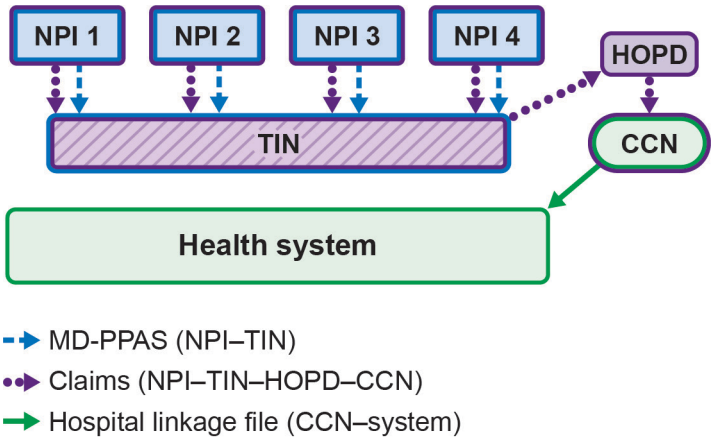
Figure III.2 illustrates how this approach links group practices to health systems using the setting of care (HOPDs in this case) for services provided by the physicians in the group practices:

- The figure shows a group practice (TIN) in the MD-PPAS v2.4 file with four physicians (NPIs).

using HOPD as the place of service. Even if the physician practice is not formally owned by the hospital, forgoing physician practice revenue by substantially billing physician outpatient services in HOPDs has been noted to imply a tight relationship between the physician practice and the hospital (or system that owns the hospital facilities).[1]

- The physicians in the TIN billed a percentage of their services (as measured by DOSs) at HOPDs to the specific HOPD in the figure (illustrated by the dotted purple lines between the NPIs and the TIN).
- The percentages for the NPIs were aggregated to the TIN level, which is illustrated by the purple shading for the TIN.
- The specific HOPD is located at the hospital (CCN) in the figure, which links the TIN to the CCN (illustrated by the dotted purple lines between the TIN and HOPD and HOPD and CCN).
- Finally, through the hospital linkage file, we connected the CCN (and the group practice) to the Compendium system (illustrated by the solid green arrow).

Figure III.2. Illustration of the hospital-based billing approach using services at HOPDs



Studies of physician-hospital integration have used similar approaches to link physicians to hospitals.[1,2] For example, one study identifies physicians as financially integrated with a hospital if they billed 90 percent or more of outpatient services in an HOPD setting.[1]

The authors excluded physicians in specialties that are primarily inpatient based when linking physicians to hospitals, such as anesthesiology, pathology, critical care, and emergency medicine.[xxvi] Their rationale is similar to the rationale for why we use billing at all hospital-based settings instead of the HOPD setting only for hospital-based specialties; that is, physicians in these specialties may bill a larger portion of their services in hospital-based settings that are not HOPDs.

Linkages based on all hospital-based billing. It is important to identify linkages between group practices with high percentages of physicians in hospital-based specialties that provide many of their services in hospital-based settings (for example, anesthesiologists and hospitalists). Therefore, we developed a version of the approach that uses services for Medicare beneficiaries billed from *all* hospital-based settings in 2021 (not just HOPDs) to assess whether these TINs are in systems.

[xxvi] Unlike the approach we used to calculate HOPD billing and develop linkages at the group practice level, the authors calculate an HOPD share at the NPI level for the purposes of their analysis.

Hospital-based settings include HOPDs, inpatient facilities, and emergency departments. We identified candidate group practice linkages to Compendium systems when most of the practice's DOSs billed at hospital-based settings occurred at specific hospitals linked to a Compendium system and at least 10 DOSs were at hospital-based settings in 2021. We applied this version of the approach based on billings at all hospital-based settings to group practices with a majority of physicians in the hospital-based specialties noted above.

Linkages made through this version of the approach are based on a broader set of services across all hospital-based settings. In addition, they are based on the assumption that group practices performing a high portion of their services in these settings all tied to the same system in 2021 likely have a tight relationship with that system. However, because the linkages are based on the set of billings provided at hospital-based settings and we only required that the TIN had 10 DOSs at hospital-based settings, it is possible that some of the TINs are not in fact part of the systems. It is also possible that some TINs in hospital-based specialties (for example, emergency medicine, pathology, or anesthesiology) contract with a hospital to provide most of their services but remain independent and not part of systems.

To minimize these types of false positive linkages, we require linkages through this approach to be corroborated by at least one of the other approaches, which are based on different connections from different data sources. Later in this section, we describe in more detail how we combined the information from the four approaches to make the final linkages.

We identified candidate linkages through the hospital-based billing approach for 5,076 group practices and 364,795 physicians. Of these linkages, 3,102 of the group practices and 320,474 of the physicians were linked through the HOPD setting; 1,974 of the group practices and 44,321 of the physicians were linked through all hospital-based settings.

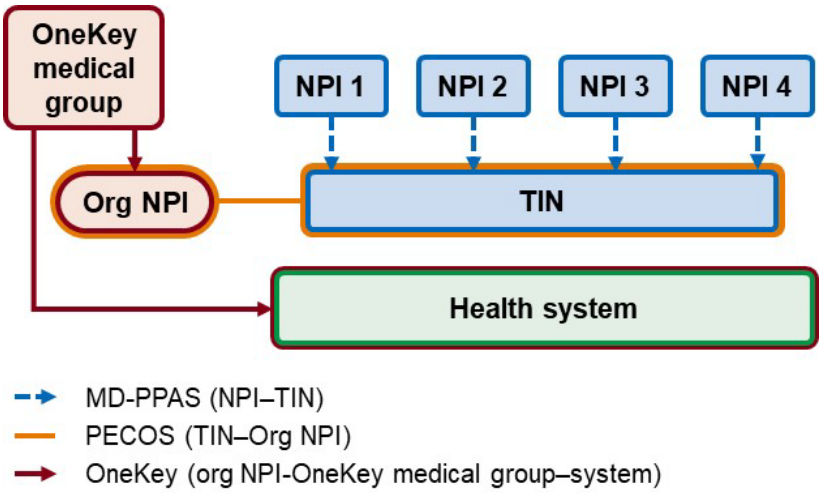
3. Organizational NPI (Org-NPI) approach

We combined the information on medical group-system assignments in the 2021 OneKey data with linkages from Org-NPIs to TINs in the 2021 PECOS data to help link group practices to systems. The common identifier in these two data sources is the Org-NPI, which is available for about 54 percent of the OneKey medical groups reported in the OneKey data.

In Figure III.3, we show an example of a group practice (TIN) in the 2021 MD-PPAS v2.4 file with four physicians (NPIs). Because the PECOS data links Org-NPIs to TINs (illustrated by the solid orange line) and the OneKey data assign medical groups with Org-NPIs to systems (illustrated by the solid red arrow), we were able to merge the two data sources by Org-NPI.[xxvii] We then linked TINs to candidate systems through this Org-NPI link; that is, if a TIN links to a OneKey medical group through their Org-NPI and the medical group is in a system (illustrated by the solid red arrow), we linked the group practice to the system through this approach.

[xxvii] Although a small percentage of TINs link to multiple Org-NPIs, as long as the Org-NPIs for a TIN link to the same system, the TIN will link the system through the Org-NPI approach.

Figure III.3. Illustration of the Org-NPI approach



In some cases, this approach linked a TIN to a system using only a small fraction of the group practice's physicians, which at times led to clearly incorrect linkages based on comparisons with linkages made through the other approaches and manual review. When we compared OneKey medical groups with TINs, many times the medical groups had many fewer physicians since they typically represent practice site locations rather than the entire group practice. Thus, a OneKey medical group might link to a TIN, but the medical group site is only part of the larger physician organization and it contains only a small fraction of the physicians in the TIN.

If no other OneKey medical group site locations in the same system linked to the TIN, then this approach linked the TIN to a system based on the small fraction of the TIN's physicians, giving us lower confidence that the group practice should be linked to the system. Thus, to provide greater confidence that the TIN linked to the correct system under this approach, we required that either the majority of the TIN's physicians or at least 50 physicians be linked to the candidate system. After applying this restriction, we linked 2,681 group practices and 267,948 physicians to Compendium systems through the Org-NPI approach.[xxviii]

4. Dominant system percentage (DSP) approach

We combined the NPI-level information on individual physician NPI-system assignments in the 2021 OneKey data with the linkages between physician NPIs and group practice TINs in the 2021 MD-PPAS v2.4 file to help link group practices to systems. Because the MD-PPAS file links TINs to individual physician NPIs and the OneKey data link individual physician NPIs to systems, we were able to merge the two data sources by NPI.

Nearly all (99.1 percent) of the physicians in the group practices found in MD-PPAS are also in the OneKey data.[xxix] This merge gave us the percentage of a group practice's physicians at the

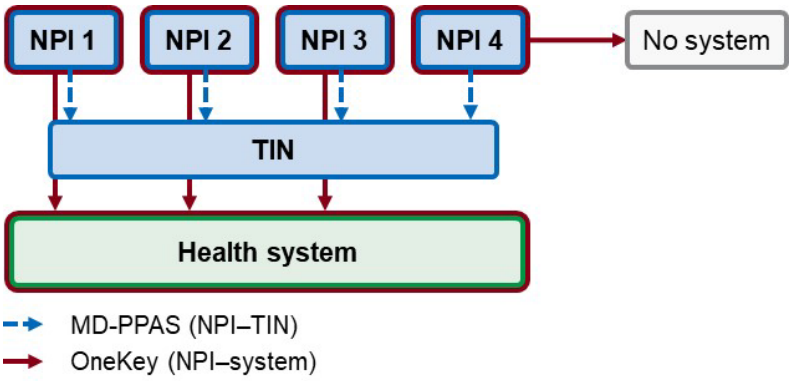
[xxviii] Before applying this restriction, we identified candidate Compendium systems for 3,210 group practices and 277,162 physicians through the Org-NPI approach.

[xxix] The OneKey data included physicians not in the subset of physicians we used from the MD-PPAS v2.4 file (for example, pediatricians not billing Medicare and physicians in solo practices, which we excluded from the MD-PPAS file). These physicians were not considered in this approach to linking group practices to systems.

TIN level (in both the MD-PPAS file and the OneKey data) that were in a system. We refer to the largest of these percentages as the dominant system percentage (DSP), and we considered the system with the DSP as a potential link to the given group practice. For example, if a group practice has 100 physicians in both data sources and the OneKey data assigned 90 in system A, 5 in system B, and 5 in no system, the DSP (system A) would equal 90 percent.

Figure III.4 illustrates how the DSP approach links group practices to health systems using 2021 OneKey data (to link NPIs to systems) and the 2021 MD-PPAS v2.4 file (to link NPIs to TINs). The figure shows an example of a group practice (TIN) in the MD-PPAS file with four physicians (NPIs). Three of the four physicians are linked to the system in the diagram through the OneKey data (illustrated by the solid red arrows); the fourth physician is not assigned to a system in the OneKey data. Since three out of four NPIs in the TIN can be linked to a specific system, the DSP for this TIN and system is 75 percent.[xxx]

Figure III.4. Illustration of the DSP approach



When calculating the percentage of a TIN's NPIs linked to systems, we excluded:

1. Physicians assigned to more than one system in the OneKey data and those linked to a system but also linked to a facility, such as a practice site, that is not affiliated with a system and
2. Physicians who switched to a different TIN during 2021 in the MD-PPAS v2.4 file.

In the OneKey data, about 5 percent of the physicians (among those also in the MD-PPAS file) were linked to more than one system, and we could not determine if the physician was more closely linked to one of these systems. In the MD-PPAS file, we defined physicians switching TINs as those consistently billing to one TIN as the primary TIN for any number of consecutive months at the beginning of 2021, then switching to consistently billing another TIN as the primary TIN for the remaining months of 2021. For example, if a physician primarily billed TIN A in months 1 through 8 and TIN B in months 9 through 12, we identified it as switching TINs during the year.

[xxx] Alternatively, if NPI 4 was not in the OneKey data, the DSP would be 100 percent because all three NPIs in both data sources (NPIs 1-3) are in the system.

We identified 23,167 physicians as switching TINs in the MD-PPAS v2.4 file. We made these exclusions so that we based the percentages on physician assignments to a single system and single TIN. We tested the sensitivity of these decisions, and they had little impact on the ultimate TIN assignment to systems.

Finally, to provide greater confidence that the group practice was linked to the correct system, we required that either the majority of the TIN's physicians or at least 50 physicians be linked to the candidate system through the approach. After applying this restriction, we linked 5,323 group practices and 367,086 physicians to Compendium systems through the DSP approach.[xxxi]

5. Summary of approaches

Table III.1 summarizes the four approaches, and Table III.2 reports the number of TINs and NPIs linked through each approach. One-quarter of group practices (8,977) and most physicians (435,370, or 70 percent) were linked through one or more approaches. The hospital-based billing and DSP approaches resulted in the most candidate linkages of group practices to systems. In the next section, we describe how we used the candidate linkages identified in these approaches to assign the final linkages between group practices and systems.

Appendix A reports all combinations of the four approaches reflected in the final set of group practice linkages.

Table III.1. Summary of approaches used to link group practices to systems

Approach	Data Sources (Linkages)	Summary	Restrictions
CCN	PECOS (TIN-CCN); hospital linkage file (CCN-system)	Group practices (TINs) linked to systems through their CCNs found in PECOS	Group practices linked to more than one system are excluded.
Hospital-based billing	Claims (NPI-TIN-CCN); hospital linkage file (CCN-system)	Group practices linked to systems based on their physicians providing services at hospital-based settings linked to systems	Majority of services at hospital-based settings are linked to the system and beneficiary dates of service are ≥10.
Org-NPI	PECOS (TIN-Org-NPI); OneKey (Org-NPI-OneKey medical group-system)	Group practices linked to systems through their Org-NPIs, and their Org-NPIs linked to systems	Majority of physicians or ≥50 physicians are linked through the approach.
DSP	MD-PPAS v2.4 (TIN-NPI); OneKey (NPI-system)	Group practices linked to systems through their physicians	Majority of physicians or ≥50 physicians are linked through the approach.

Key: CCN = CMS Certification Number; PECOS = Provider Enrollment, Chain, and Ownership System; TIN = Tax Identification Number; NPI = National Provider Identifier; Org-NPI = organizational NPI; DSP = dominant system percentage; MD-PPAS = Medicare Data on Provider Practice and Specialty.

[xxxi] Before applying this restriction, we identified candidate Compendium systems for 11,132 TINs and 492,563 physicians through the DSP approach.

Table III.2. Number of group practices linked to Compendium systems through each approach

Approach	Number of Group Practices	Number of Physicians
All group practices and physicians	**36,202**	**621,717**
Not linked through any approach	27,225	186,347
One or more approaches	8,977	435,370
CCN	1,450	83,786
Hospital-based billing	5,076	364,795
Org-NPI	2,681	267,948
DSP	5,323	367,086

Key: CCN = CMS Certification Number; Org-NPI = organizational NPI; DSP = dominant system percentage.

Note: The counts of practices and physicians linked to Compendium systems through each of the four approaches (CCN, hospital-based billing, Org-NPI, DSP) will not sum to the counts in the "One or more approaches" line because many practices are linked to a system through more than one of the approaches, and therefore, are included in more than one of the counts for the four approaches. For a complete breakdown of counts for each combination of approaches through which group practices were linked, please refer to Appendix A.

B. Decision Rules for Assigning Group Practices to Systems

This section summarizes the decision rules we used to link group practices to systems based on information gleaned from the four approaches described above. With the exception of group practices linked to a system through the CCN approach, we required that linkages be based on more than one approach to accept the linkage.

We accepted all linkages based on the CCN approach because the linkages are based on only two connections and we have relatively high confidence in both. The connections between the TIN and CCN came directly from Medicare enrollment data, and the connections between the CCN and system came from the hospital linkage file.

Linkages made through the other three approaches rely on connections between physicians or OneKey medical groups to systems through the OneKey data or indirect connections surmised through aggregate billing of a TIN's physicians. To be conservative, we required two or more corroborating approaches linking the group practice to the same system.

Finally, we used name matching or manual review to adjudicate a small number of cases in which the approaches linked to different systems or for large TINs with a link through only one of the approaches (excluding the CCN approach).

The decision rules followed four steps:

1. We accepted group practice linkages to a system through the CCN approach.
2. We accepted group practice linkages to a system through two or more of the remaining three approaches.
3. We examined whether group practice names matched system names to resolve discrepancies (group practices linked to different systems through two or more approaches) or to corroborate linkages made through only one of the approaches (excluding the CCN approach).

4. We conducted manual review using web searches of 19 large TINs (with 50 or more physicians each) that still linked to more than one health system to identify the correct system linkage.

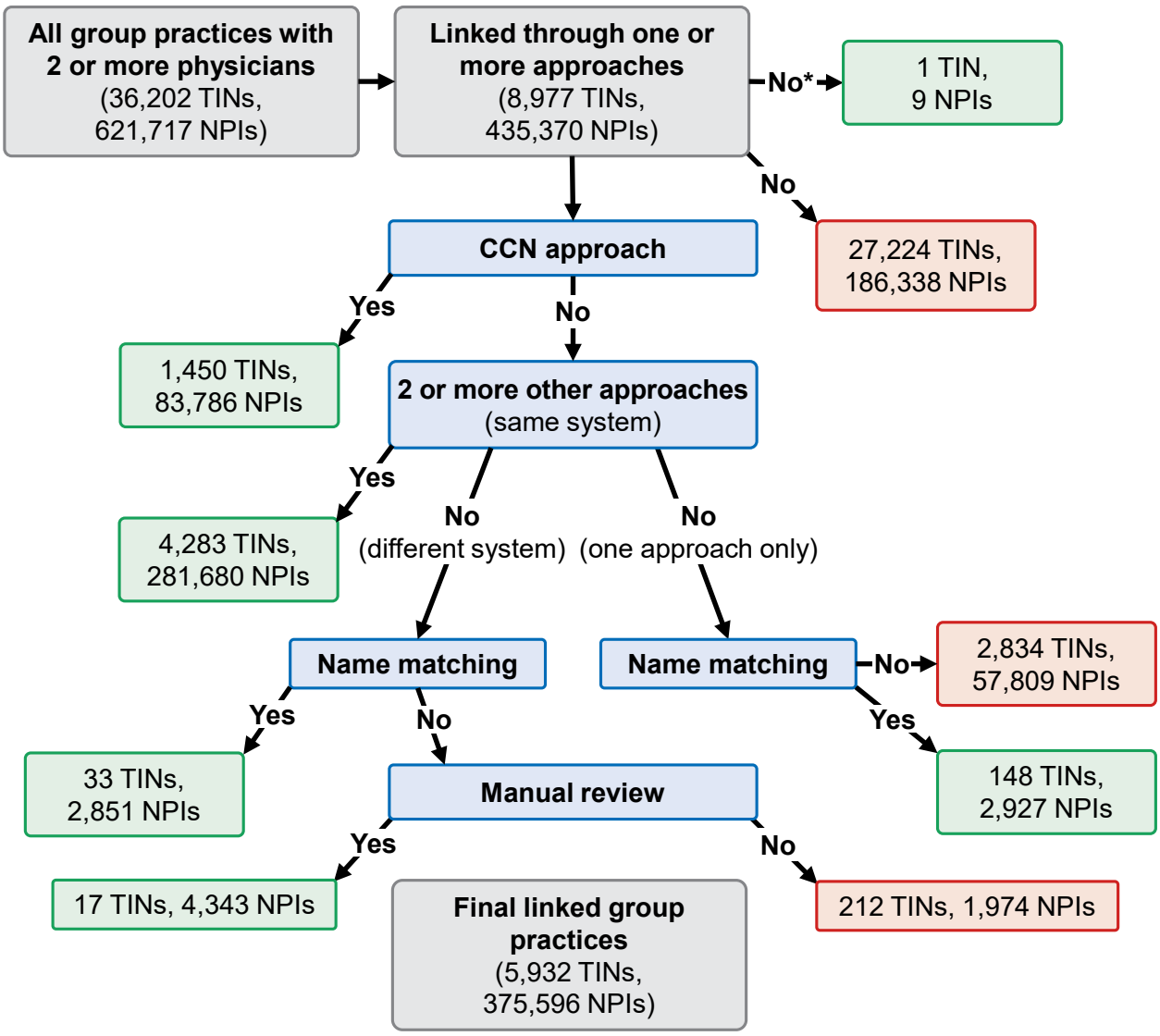
We also added linkages for four group practices based on manual review of possible linkages to five systems with no group practices based on the four steps in the process.

Figure III.5 illustrates the application of the decision rules in the four steps to identify the group practice-system linkages. The figure also shows the number of group practices and physicians linked in the steps.

We began with 36,202 group practice TINs with 621,717 physician NPIs in the 2021 MD-PPAS v2.4 file (restricted to TINs with two or more physicians). Of these group practices, 8,977 (435,370 physicians) linked to a system through one or more of the four approaches.

We found that 27,224 group practices with 186,338 physicians did not link to a Compendium system through any of the approaches and are reported as not being part of a Compendium system in the final group practice linkage file. We then applied the four steps to use the information in these candidate linkages to identify the linkages to Compendium systems.

Figure III.5. Decision rules for linking group practices to systems



Note: "Yes" indicates that the group practices were linked through the approach and are reflected as being in systems in the final group practice linkage file; "No" means that they were not linked and are reflected as not being in systems in the file. "No*" reflects the single group practice that was not linked using any of the four approaches, but was ultimately linked to a system in the final group practice linkage file through a manual review of systems with no linked group practices.

1. Linking group practices through the CCN approach

We accepted all linkages made between group practices and Compendium systems made through the CCN approach, except for cases in which the approach linked the group practice to multiple systems. (Our procedures for handling cases linked to multiple systems in general are described under steps 3 and 4 below.) This first step resulted in 1,450 group practices with 83,786 physicians linked to Compendium systems.

2. Linking group practices with two or more approaches

Next, we accepted group practices linked by two or more approaches (DSP, hospital-based billing, or Org-NPI) if all the available approaches linked to the same system. In this step, we linked 4,283 group practices with 281,680 physicians to Compendium systems (Figure III.5). Among these linkages, the most common combination of approaches relied on the DSP and hospital-based billing approaches (1,426 group practices with 30,383 physicians). The combination that led to the linkage of the largest number of physicians was DSP, Org-NPI, and HOPD approaches (1,259 group practices with 190,263 physicians). Appendix A reports all combinations of the four approaches reflected in the final set of group practice linkages.

3. Name matching

We used name matching between the TIN names (from the MD-PPAS v2.4 file) and health system names (from the systems list) to resolve discrepancies in which TINs were linked to different systems through the approaches (excluding linkages through the CCN approach). In other words, name matching served to break ties.

We used two basic approaches to identify matches:

- We used the COMPGED function in SAS to assign match scores to each TIN and candidate system pair; and
- We identified whether the first two or more words of the names matched identically.

We accepted matches with a COMPGED score of 150 or lower (lower scores reflect closer matches), which reflects an extremely close match.[xxxii] We also accepted TIN-system name pairs in which the first two words matched identically.

Based on a review of the candidate matches, we have a high degree of confidence that pairs meeting these criteria are, in fact, group practices owned or tightly managed by the matched systems. We identified linkages for 33 group practices with 2,851 physicians through name matching (Figure III.5).

We also used name matching to corroborate linkages for TINs in which only one of the approaches linked the group practice to a system. We identified linkages for 148 group practices with 2,927 physicians through this use of name matching. Conversely, 2,834 group practices with 57,809 physicians were linked through only one approach and did not have a name match. Taken together, name matching adjudicated or confirmed linkages for 181 group practices with 5,778 physicians.

4. Manual review

After completing the three steps to link group practices described above, we manually reviewed 19 group practices with 50 or more physicians that linked to different systems through two or

[xxxii] We manually reviewed a sample of the name matches to ensure that this approach generally linked group practices to the correct system. All the cases reviewed linked to the correct systems.

more approaches but could not be resolved by name matching.[xxxiii,xxxiv] Two researchers conducted web searches of the group practices and candidate systems to determine which of the systems (if any) owned or tightly managed the group practices. Specifically, they reviewed the official web pages of the group practices and systems (e.g., "about us" pages and list of practice locations), as well as other information reporting that the two were linked (e.g., news articles announcing mergers that would link the group practice to a system).

The review also considered shared branding on the websites as evidence that the group practice and system should be linked. We compared the linkages made by the two researchers and met with a third researcher to resolve any discrepancies.

Through the review, we linked an additional 14 group practices (of the 19 reviewed) with 4,319 physicians to systems. Appendix B lists the 14 group practices and their systems.

Before we finalized the linkages, we compared the number of group practices and physicians in each of the 635 Compendium systems with the totals based on the linkage file. We found linkages for 630 of the 635 systems (99.2 percent), with only 5 systems having no linked TINs.

Because each Compendium system was required to have at least one group practice, we further reviewed possible linkages to these five systems. More specifically, we reviewed all group practices with any evidence from one of the four approaches indicating that it could be linked to one of these candidate systems. For example, we reviewed group practices linked to these candidate systems through one of the approaches (excluding the CCN approach) that lacked corroboration from a second approach.

We also reviewed group practices that had an association with one of these candidate systems but did not meet the criteria's threshold for the approach; for example, a group practice associated with a system through the DSP approach but with less than 50 percent of the practice's physicians linked through the approach. Based on this review, we identified group practices associated with three of the five Compendium systems with no linked TINs.

Using the same approach as the manual review described above, two researchers conducted web searches of the group practices and candidate systems to determine if the group practices should be linked to the systems. We compared the linkages made by the two researchers and met to resolve any discrepancies. We determined that four of the group practices (33 physicians) should be linked to systems, which led to three additional Compendium systems with at least one group

[xxxiii] We selected 50 as the cutoff because (1) there was a large dropoff in the number of physicians below 50, with only one group practice having 40-50 physicians and the vast majority having fewer than 10; (2) 50 is the number of physicians used in the Compendium definition of a system, which means that we are reviewing all group practices large enough to result in a linked system meeting the definition; and (3) this cutoff resulted in a number of group practices that could feasibly be reviewed.

[xxxiv] There were 207 group practices with fewer than 50 physicians each that were linked to different systems and had no name match. We did not further attempt to match these group practices to systems; they are classified as not being part of systems in the group practice linkage file.

practice; we list the group practice names and systems in Appendix B.[xxxv] One of these four group practices was not linked using any of the four approaches.

5. Final group practice linkage file

Table III.3 reports the number of group practices and physicians linked through each step and in total in the final group practice linkage file. We linked 5,932 group practices with 375,596 physicians to Compendium systems. Thus, 16.4 percent of the 36,202 group practices (with two or more physicians) and 60.4 percent of the 621,717 physicians in these group practices are linked to Compendium systems.

Table III.3 also reports analogous figures from the 2020 group practice linkage file. The percentage of group practices in systems in 2021 is the same as in 2020, but the percentage of physicians in systems in 2021 is higher than in 2020, when 6,094 of 37,109 group practices (16.4 percent) and 361,718 of 612,163 physicians (59.1 percent) were linked to systems.[xxxvi]

In addition, we report the comparison of 2020 and 2021 linkages overall and by approach in Appendix A. The numbers of group practices and physicians linked to systems generally remained similar for most combinations of approaches.

Table III.3. Number of group practices linked to systems in 2020 and 2021

	2021		2020	
Step in the Process	**Number of Group Practices**	**Number of Physicians**	**Number of Group Practices**	**Number of Physicians**
Linkages through CCN approach	1,450	83,786	1,474	82,469
Linkages through two or more other approaches	4,283	281,680	4,408	269,279
Name matching				
To adjudicate discrepancies	33	2,851	35	2,916
To corroborate linkage through a single approach	148	2,927	162	2,926
Manual review to adjudicate additional discrepancies	14	4,319	12	4,045
Manual review of possible linkages to systems with no linked group practices through prior steps	4	33	3	83
Total in systems	**5,932**	**375,596**	**6,094**	**361,718**

[xxxv] The two remaining Compendium systems with zero group practices linked are Alecto Healthcare Services (HSI00000021) and Childrens Hospital and Health System (HSI00000212).

[xxxvi] Fewer group practices had two or more physicians in 2021 than in 2020 (36,202 compared with 37,109), but these group practices had more physicians (621,717 compared with 612,163). There were 33,155 group practices in 2020 and 2021 (3,954 in only 2020 and 3,047 in only 2021). Of the 33,155 TINs in 2020 and 2021, 5,019 are linked to systems in both years, 26,850 are not linked to a system in both years, 616 are linked to a system in 2020 but not in 2021, and 670 are linked to a system in 2021 but not in 2020. Of the 616 group practices linked to systems in 2020 but not in 2021, many are hospital-based specialty groups whose affiliations in 2021 have expanded beyond one system and thus are less obviously aligned with the prior system.

IV. Contents of the Group Practice Linkage File

A. Variables Included in the Group Practice Linkage File

Table IV.1 contains the variables included in the group practice linkage file. The file includes two variables that can be used to directly identify group practices: (1) the TIN legal name, taken from a combination of the 2021 fee-for-service MD-PPAS v2.4 file, MA data, and PECOS; and (2) the PECOS Associate Control ID (PAC ID), which is assigned in PECOS to uniquely identify Medicare enrollments.[xxxvii]

For group practices linked to Compendium systems, the linkage file also contains the unique Compendium health system IDs and system names, which link directly to the Compendium of U.S. Health Systems. In addition, the linkage file includes five variables taken directly from the 2021 MD-PPAS v2.4 file that can be used to link the data to the official fee-for-service MD-PPAS data:

1. TIN name from MD-PPAS,
2. State where the plurality of the TIN's NPIs are located,
3. Total number of physicians in the TIN,
4. Total number of nurse practitioners and physician assistants in the TIN, and
5. Total number of line items billed through the TIN.[xxxviii]

The second TIN name variable overlaps substantially with the first TIN name on the linkage file, except that it only includes values from the fee-for-service MD-PPAS file; that is, it does not include TIN names from the MA data or PECOS when MD-PPAS is missing the TIN name.

Finally, the file includes a flag that identifies the TINs in the 2021 group practice linkage file found in the 2021 MA data but not the fee-for-service MD-PPAS file and thus will not link to the official 2021 MD-PPAS data. AHRQ added this set of variables to enable users to link the file to the 2021 MD-PPAS data.

Appendix C contains a data dictionary for the linkage file. Appendix D provides a step-by-step summary of how users can link the group practice linkage file to the MD-PPAS data.

[xxxvii] We obtained the TIN name from Medicare Advantage data or PECOS when the name was missing from the 2021 fee-for-service MD-PPAS v2.4 file. The group practice linkage file does not include the actual TIN of the group practice, because this information is not publicly available elsewhere and thus cannot be included in the file.
[xxxviii] A small number of TINs had zero or one reported physician based on information in the 2021 fee-for-service MD-PPAS data. It is possible for a TIN to have fewer than two physicians in the fee-for-service MD-PPAS data because we supplement these data with Medicare Advantage data (see section II.A); thus, these TINs have two or more physicians in the combined Medicare Advantage and fee-for-service MD-PPAS v2.4 file.

Table IV.1. Variables included in the group practice linkage file

Variable	Description	Source
TIN name	Group practice legal name	MD-PPAS; Medicare Advantage data; PECOS
PAC ID	PECOS Associate Control ID, used to uniquely identify Medicare enrollments, in this case, the group practices (TINs)	PECOS
Health system ID	Compendium health system ID	Compendium of U.S. Health Systems
Health system name	Compendium health system name	Compendium of U.S. Health Systems
Variables from MD-PPAS to be used to link directly to the 2021 MD-PPAS data		
TIN name	Group practice legal name	MD-PPAS
State	State in which the plurality of the group practice's NPIs are located	MD-PPAS
Number of physicians	Total number of physicians in the TIN (based on primary TIN assignment)	MD-PPAS
Number of nurse practitioners and physician assistants	Total number of nurse practitioners and physician assistants in the TIN	MD-PPAS
Number of line items	Total number of line items from Medicare claims billed through the TIN	MD-PPAS
MA-only TIN	A flag indicating that the TIN is only found in the 2022 Medicare Advantage data, not in the 2022 fee-for-service MD-PPAS data; only TINs with a value of zero for this variable will link to the MD-PPAS data.	Medicare Advantage data

B. Linking the Group Practice Linkage File to Other Data Sources

Users of the group practice linkage file can link the data to other data sources through the TIN organization name or PAC ID. For example, users can link the data using PAC ID to files derived from PECOS, such as the Public Provider Enrollment data or the Physician Compare National Downloadable File.[xxxix]

Through any of the linkages described in this section, users can identify the physicians in the group practices and thereby the physicians tightly affiliated with the Compendium health systems. In turn, users can link information in Medicare claims data to systems using physician NPIs. By making such linkages possible, the group practice linkage file enables users to examine a wide variety of issues related to the relationships between group practices (and their physicians) and systems and how group practices in systems compare with those not in systems.

[xxxix] These data are available at https://data.cms.gov/public-provider-enrollment and https://data.medicare.gov/data/physician-compare.

1. Using the TIN name to link TINs to other data sources

The vast majority (98.8 percent) of group practices in the linkage file have unique TIN names (Table IV.2). Users can link the data to other data sources that use the TIN names from the MD-PPAS data, such as data from the Internal Revenue Service's 990 forms, which include information on nonprofit hospitals.

Table IV.2. Summary of unique and repeated TIN names

Categories of group practices	Number of Group Practices	Percentage of Group Practices
All group practices in the linkage file		
With unique names	35,769	98.8%
With names repeated one or more times	433	1.2%
Group practices in health systems		
With unique names	5,711	97.3%
With names repeated one or more times	161	2.7%
Group practices not in health systems		
With unique names	29,998	99.1%
With names repeated one or more times	272	0.9%

2. Using the PAC ID to link TINs to other data sources

PAC ID is a unique identifier assigned by PECOS to identify Medicare enrollments. PAC IDs have an almost one-to-one relationship with group practices; 0.1 percent of group practices (41 of 36,202) link to more than one PAC ID. PAC ID can be used to link the group practice linkage file to files derived from PECOS, such as the Public Provider Enrollment data, which can then be used to link the TIN and other PECOS enrollments to the group practice and Compendium systems. The PAC ID can also be used to link the group practice linkage file to the Physician Compare National Downloadable File, which can be used to link physicians and performance information to the group practices and Compendium systems.

C. Comparison of Physician Counts in the Group Practice Linkage File and 2021 Systems List

The 2021 Compendium of U.S. Health Systems contains aggregate counts of all physicians and primary care physicians. As described in detail in the technical documentation for the systems list, the counts represent the highest counts across the key data sources (OneKey and AHA) used to generate the 2021 systems list. These counts varied substantially between the two data sources, representing both differences in the physicians included in the counts and the underlying approaches used by the data sources to identify physicians linked to systems.

Table IV.3 contains information on physician counts reported in the 2021 systems list and in the group practice linkage file. The latter are calculated as the sum of physicians in the group practices linked to the systems. The count of physicians for a group practice is calculated as the physicians in the MD-PPAS v2.4 file combined with the MA data (that is, those billing Medicare Part B or Medicare Advantage) for which the group practice is the primary TIN for the physician.

Table IV.3. Comparison of system-level physician counts

Type of physicians by data file	Number of Systems	Physicians Linked to Systems		
		Mean	Minimum	Maximum
Total physicians				
Systems list	635	977	50	24,888
Group practice linkage file	633	593	4	22,275
Difference	2	383	46	2,613
Primary care physicians				
Systems list	633	321	13	11,197
Group practice linkage file	628	157	1	9,039
Difference	5	164	2	2,158

Note: The number of systems listed is the number of systems with at least one physician (or primary care physician) reported in the 2021 systems list or linked through a group practice in the 2021 group practice linkage file. For example, all but 2 of the 635 systems have at least one group practice and two physicians linked to them in the group practice linkage file. The number of systems with a primary care physician based on the counts in the systems list is 633 because two systems are only found in the AHA data, which does not include counts of primary care physicians (only total physicians). The specialties indicating primary care are adolescent medicine, family medicine, geriatrics, general practice, internal medicine, and pediatrics.

All but 2 of the 635 Compendium systems had at least one group practice assigned to them in the group practice linkage file. Furthermore, 551 systems (86.8 percent) would meet the inclusion criteria for a Compendium system (that is, 50 total physicians and 10 primary care physicians) if we used the group practice linkage file physician counts (results not reported).

However, when comparing physician counts at the system level, we found some notable differences between the physician counts in the systems list and the group practice linkage file counts. Specifically, we identified fewer total physicians and primary care physicians per system in the group practice linkage file. The average number of physicians in a system based on the systems list count was 977 compared with 593 based on the group practice linkage file. In addition, some systems had fairly large differences (for example, one system has 14,179 physicians reported in the systems list but only 5,747 physicians based on the group practice linkage file).

While most systems had higher physician counts based on the Compendium, a small number of systems had higher counts based on the group practice linkage file. For example, one system had 1,561 physicians reported in the group practice linkage file but only 1,359 physicians based on the systems list.

Since the total number of physicians reflected in the group practice linkage file is much lower than totals reported in the systems list (621,717 compared with nearly 1.1 million physicians), differences between the systems' physician counts are to be expected. Some of the difference in physician counts is due to the group linkage file being based on physicians in the MD-PPAS v2.4 file, which only includes physicians billing Medicare Part B, and the restriction to group practices with two or more physicians (thus excluding physicians in solo practices).

Additionally, some of the difference is due to how physicians count towards systems in the systems list versus in the group practice linkage file. In the systems list, when a physician is

linked to more than one system, the physician is included in the counts for all of their systems. In the group practice linkage file, physicians can only be linked to one system, based on the primary TIN designation in the MD-PPAS v2.4 file.

Finally, the counts of primary care physicians in the systems list could be higher because they include physicians who focus primarily on the care of hospitalized patients (hospitalists). For the systems list, because the data sources used to determine systems' physician counts do not indicate the setting, the counts for primary care physicians include some hospitalists. In contrast, the source for physicians in the group practice linkage file, the MD-PPAS v2.4 file, identifies hospitalists, which enables us to remove them from counts of primary care physicians.

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

V. Caveats and Limitations

We note several caveats and limitations related to the methods used to create the group practice linkage file, its contents, and uses of the file. First, the group linkage file only includes group practices with physicians in the 2021 MD-PPAS file; that is, physicians billing Medicare Part B in 2021. Similarly, the file only includes group practices with two or more physicians (although physicians in solo practice can also be identified in MD-PPAS). Thus, the group practice file does not represent linkages to systems for solo practices or physicians and practices that do not bill Medicare Part B (for example, many pediatric practices). However, the file represents the majority of physicians practicing in the United States in 2021. For example, compared with the count of physicians represented in the systems list, which includes all active physicians in 2021, the group practice linkage file reflects nearly 60 percent of physicians.

Although this file does not include information on system affiliation for solo practices, anecdotal evidence suggests it is unlikely that many solo practices identified in the MD-PPAS v2.4 file were part of systems. When purchasing solo practices, systems typically incorporate these solo physicians into other larger TINs that are part of the system. Thus, these formerly solo physicians would likely be in a larger TIN and thus counted as part of the system in the group linkage file. Nonetheless, it is important for users to consider the composition of the linkage file when linking it to other data sources and how it could affect the planned analysis.

Although we required corroborating evidence to link group practices to systems (except for the CCN approach), it is still possible that we have mistakenly linked some group practices to systems (that is, false positives). It could be that two of the approaches mistakenly linked a group practice to the same system; for example, the DSP and Org-NPI approaches both rely in part on linkages made in the OneKey data (NPIs and medical group locations, respectively). The linkages may be incorrect or out of date, or they may not quite reflect the ownership or tight management relationships intended by the linkage file.

It is also possible that a group practice could be mistakenly linked to a system through the CCN approach if a hospital (CCN) linkage to the system was incorrect. Similarly, the thresholds we use for the various approaches generally require most physicians or beneficiary dates of service to link to a given system. However, these thresholds could be too lenient, which would lead to too many linkages. Conversely, the thresholds could be too strict, which would lead to too few linkages. Ultimately, we chose thresholds that would identify defensible linkages for the vast majority of group practices in systems, and we required linkages to be confirmed through another approach.

In the hospital-based billing approach, inpatient and emergency department services may not be a good indicator of system ownership or tight management for cases in which services are provided by independent medical staff or physicians are employed by a large national company. In addition, the hospital-based billing approach could create false positives for group practices with percentages measured by relatively few services provided through hospital-based settings. To minimize false positives, we required evidence from two or more approaches to be more confident regarding the group practices assigned to systems.

We acknowledge that the requirement to confirm linkages through another approach may result in missing some number of group practices that should be assigned to systems (false negatives). For example, the hospital-based billing approach may not identify group practices that are owned but not billing through an HOPD (such as off-campus practices). Because users can link the file to other data sources, they can make further decisions regarding whether group practices in the data should be linked to systems.

References

1. Neprash HT, Chernew ME, McWilliams JM. Little evidence exists to support the expectation that providers would consolidate to enter new payment models. Health Aff (Millwood). 2017;36(2):346-354. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5458639/. Accessed October 30, 2024.
2. Song Z, Wallace J, Neprash HT, McKellar MR, Chernew ME, McWilliams JM. Medicare fee cuts and cardiologist-hospital integration. JAMA Intern Med. 2015;175(7):1229-1231. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4664603/. Accessed October 30, 2024.
3. Furukawa MF, Machta RM, Barrett KA, Jones DJ, Shortell SM, Scanlon DP, Lewis VA, O'Malley AJ, Meara ER, Rich EC. Landscape of health systems in the United States. Med Care Res Rev. 2019 Jan 23. https://doi.org/10.1177/1077558718823130. Accessed October 24, 2024.
4. Cohen GR, Jones DJ, Heeringa J, Barrett K, Furukawa MF, Miller D, Mutti A, Reschovsky JD, Machta R, Shortell S, Fraze T, Rich E. Leveraging diverse data sources to identify and describe U.S. health care delivery systems. eGEMs. 2017 Dec;5(3):9. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5983023/. Accessed October 30, 2024.
5. Report to the Congress: Medicare and the Health Care Delivery System. Washington, DC: Medicare Payment Advisory Commission; June 2013. https://www.medpac.gov/document/june-2013-report-to-the-congress-medicare-and-the-health-care-delivery-system-12/. Accessed October 30, 2024.
6. Chatrath S, Rich EC, O'Malley AS, Cohen G, Jones DJ. Hospital outpatient department billing is a poor indicator of primary care practice integration with hospital systems. Health Serv Res. 2024 Apr;59(2):e14284. https://pubmed.ncbi.nlm.nih.gov/38287519/. Accessed November 6, 2024.

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

Appendix A. Combinations of the Four Approaches Used in the Final Set of Group Practice Linkages in 2020 and 2021

Approaches	2021		2020	
	Group Practice Count	Physician Count	Group Practice Count	Physician Count
CCN	139	915	141	1,017
CCN, DSP	123	957	124	944
CCN, DSP, HOPD	241	14,515	255	15,363
CCN, DSP, All hospital-based settings	107	2,330	99	2,186
CCN, HOPD	111	1,306	119	1,701
CCN, All hospital-based settings	51	444	61	616
CCN, Org-NPI	17	146	19	120
CCN, Org-NPI, DSP	161	2,989	165	2,147
CCN, Org-NPI, DSP, HOPD	409	56,886	394	54,538
CCN, Org-NPI, DSP, All hospital-based settings	59	2,121	60	1,651
CCN, Org-NPI, HOPD	16	528	22	525
CCN, Org-NPI, All hospital-based settings	9	138	8	55
DSP	41	786	46	701
DSP, HOPD	913	53,940	969	55,603
DSP, All hospital-based settings	1,426	30,383	1,468	28,663
HOPD	33	717	47	823
All hospital-based settings	65	1,346	60	1,392
Org-NPI	12	102	12	93
Org-NPI, DSP	362	4,897	339	3,806
Org-NPI, DSP, HOPD	1,259	190,263	1,297	180,942
Org-NPI, DSP, All hospital-based settings	222	7,019	226	5,964
Org-NPI, HOPD	120	2,319	136	2,397
Org-NPI, All hospital-based settings	35	540	27	471
None (linked through manual review only)	1	9	0	0
Total	5,932	375,596	6,094	361,718

CCN = CMS Certification Number; DSP = dominant system percentage; HOPD = hospital outpatient department; Org-NPI = organizational NPI.

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

Appendix B. TINs Assigned Through Manual Review

TIN Name	Assigned Health System
Group practices originally linked to different systems	
Blue Ridge Healthcare Medical Group Inc	HSI00001123/UNC Health Care System
Board of Regents of the University of Oklahoma OU Physicians Tulsa	HSI00001383/OU Health
Coastal Medical Physicians Inc	HSI00000585/Lifespan
Colorado Permanente Medical Group PC	HSI00000536/Kaiser Permanente
Dreyer Clinic Inc	HSI00001277/Advocate Aurora Health
ETMC Physician Group, Inc.	HSI00000046/Ardent Health Services
Harbor Medical Associates Inc.	HSI00000813/Mass General Brigham
Healthcare Partners Medical Group Coats LTD	HSI00000516/Intermountain Healthcare
Louisiana State University School of Medicine in New Orleans Faculty G	HSI00000559/LSU Healthcare Services Division
Mercy Clinics Inc	HSI00001106/Trinity Health
The Southeast Permanente Medical Group	HSI00000536/Kaiser Permanente
UT Physicians (PAC ID: 2961732193)	HSI00001165/The University of Texas System
UT Physicians (PAC ID: 8426960360)	HSI00001165/The University of Texas System
UT Physicians Specialty Services	HSI00001165/The University of Texas System
Group practices linked to systems previously with no group practices	
Advanced Multispecialty Medical Services LLC	HSI00001450/Larkin Health System
Bradley B. Bailey, MD and Roger B. Schechter, MD, Inc.	HSI00000809/Palomar Health
Larkin Emergency Physicians LLC	HSI00001450/Larkin Health System
Nebraska Pediatric Practice Inc	HSI00000213/Childrens Hospital and Medical Center

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

Appendix C. Data Dictionary

Variable Name	Variable Type	Description
tin_name	Character	TIN name from MD-PPAS (filled in with TIN name from Medicare Advantage data or PECOS if missing in MD-PPAS)
pecos_pac_ids	Character	List of PECOS Associate Control IDs assigned by PECOS; IDs are separated by a "^"
health_sys_id	Character	Unique system ID
health_sys_name	Character	Health system name
tin_name_md_ppas	Character	TIN name from MD-PPAS
state_md_ppas	Character	State in which the plurality of the group practice's NPIs are located
md_do_md_ppas	Numeric	Total number of physicians in the TIN (based on primary TIN assignment)
np_pa_md_ppas	Numeric	Total number of nurse practitioners and physician assistants in the TIN
service_lines_md_ppas	Numeric	Total number of line items from Medicare claims billed through the TIN
ma_only_tin	Numeric	A flag indicating that the TIN is only found in the 2020 Medicare Advantage data, not in the 2020 fee-for-service MD-PPAS data; only TINs with a value of zero for this variable will link to the MD-PPAS data.

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

Appendix D. Linking the Group Practice Linkage File to 2021 MD-PPAS Data

Linking the group practice linkage file with the 2021 MD-PPAS data requires (1) processing the 2021 MD-PPAS data to a primary TIN level to create the set of five variables that uniquely identify a TIN; and (2) merging to the group practice linkage file by those five variables. We describe the steps and provide sample SAS code for merging the two files.

Step 1: Processing the 2021 MD-PPAS data

1.1. Identify the most common State within a TIN.

For each TIN, we counted the number of NPIs within each State. Most TINs will include NPIs in a single State; however, some TINs have NPIs located in more than one State. We assigned each TIN the State reported most frequently for the NPIs in the TIN. For TINs with multiple States and the same number of NPIs in more than one State, we selected the first State by alphabetical order. The SAS code for this step is:

```
proc sql;
  create table count_npi_state as
  select tin1 as tin
        ,state as state_md_ppas
       ,count(*) as number_of_npi
  from {insert 2021 MD-PPAS filename}
  group by tin1
          ,state;
quit;

proc sort data=count_npi_state;
  by tin descending number_of_npi state_md_ppas;
run;

data tin_state;
  set count_npi_state;
  by tin descending number_of_npi state_md_ppas;
  if first.tin;
run;
```

1.2. Identify primary TIN name, counts of TIN specialties, and line items billed.

For each TIN, we selected the primary TIN name and created the following count variables:

- The number of NPIs with specialty codes indicating Physician (broad specialty not equal to 7 [Non-Physician] or 9 [Specialty Unknown])
- The number of NPIs with primary specialty codes indicating Nurse Practitioner (50) or Physician Assistant (97)
- The number of line items billed to TIN by all NPIs within the TIN

Some TINs in the MD-PPAS data are missing primary TIN Name; however, with the combination of the most common State, the counts of physicians, nurse practitioners, and

physician assistants, and the counts of lines billed, the TIN will merge to a unique TIN in the group practice linkage file. The SAS code for this step is:

```
proc sql;
  create table tin_name_counts as
  select tin1 as tin
        ,tin1_legal_name as tin_name_md_ppas
        ,sum(case when spec_broad ~in (7,9) then 1 else 0 end) as
md_do_md_ppas
        ,sum(case when spec_prim_1 in ("50","97") then 1 else 0 end)
as np_pa_md_ppas
        ,sum(npi_srvc_lines) as service_lines_md_ppas
  from {insert 2021 MD-PPAS filename}
  group by tin1
          ,tin1_legal_name;
quit;
```

1.3. Merge the TIN files to create a TIN-level MD-PPAS file with the five variables needed to merge with the group practice linkage file.

Merge the two TIN-level files created in steps 1.1 and 1.2 by primary TIN. The merged file will include the following variables and can be merged uniquely to the group practice linkage file:

- TIN
- STATE_MD_PPAS
- TIN_NAME_MD_PPAS
- MD_DO_MD_PPAS
- NP_PA_MD_PPAS
- SERVICE_LINES_MD_PPAS
- NUMBER_OF_NPI – variable not needed for merging to the group practice linkage file

The SAS code for this step is:

```
data md_ppas_tins;
  merge tin_state
        tin_name_counts;
  by tin;
run;
```

Step 2: Merging to the Group Practice Linkage File by the Five Variables

2. Merge the TIN-level MD-PPAS file with the group practice linkage file.

Merge the TIN-level MD-PPAS file created in steps 1.1-1.3 to the group practice linkage file after limiting the group practice linkage file to TINs with at least one NPI with fee-for-service beneficiaries (MA_ONLY_TIN = 0). This will be a one-to-one merge, in which each record in the group practice linkage file merges to one TIN in the MD-PPAS data.

The SAS code for this step is:

```
data md_ppas_gplf;
  merge md_ppas_tins (in=mdppas)
        {insert group practice linkage file filename} (in=gplf
where=(ma_only_tin = 0));
  by tin
     state_md_ppas
     tin_name_md_ppas
     MD_DO_md_ppas
     NP_PA_md_ppas
     service_lines_md_ppas;

  if mdppas and gplf;
run;
```

AHRQ Pub. No. 25-0011-5-EF
December 2024