

Comparative Health System Performance Initiative: Compendium of U.S. Health Systems, 2016, 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: HHS A290201600001C

Prepared by:

Mathematica, Washington, DC
David Jones
Rachel Machta
Eugene Rich
Ken Peckham
Sandi Nelson

AHRQ Publication No. 19(21)-0083-EF
Updated December 2020 (original version, Publication No. 19-0083, October 2019)



Acknowledgments

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

- AHRQ (Mike Furukawa, Dan Miller, Jing Guo, Linda Bergofsky, Zeynal Karaca, and Herb Wong)
- CHSP Initiative Centers of Excellence (Dartmouth, NBER, RAND)
- CHSP Technical Expert Panel
- Data vendors (IQVIA and American Hospital Association)
- Additional Mathematica staff (Jim Reschovsky, Linda Molinari, Miaomiao Shen, Gwyn Olson, and Dorothy Bellow)

Contents

Acknowledgments.....	ii
Updates to This Version.....	v
Executive Summary	1
A. Methodology.....	1
B. Using the Linkage File To Link Group Practices to Other Data Sources.....	5
I. Introduction.....	7
II. Data Sources	8
A. Identifying Group Practices and Physicians Using MD-PPAS	8
B. Other Data Sources Used To Develop Linkages	10
1. Hospital linkage file.....	10
2. Medicare Provider Enrollment, Chain, and Ownership System (PECOS).....	11
3. Medicare claims	11
4. Healthcare Organization Services (HCOS) database.....	13
III. Methodology.....	13
A. Approaches to Linking Group Practices to Systems.....	14
1. CMS Certification Number (CCN) approach	14
2. Hospital-based billing approach	15
3. Organizational NPI (Org-NPI) approach	18
4. Dominant system percentage (DSP) approach	20
5. Summary of approaches.....	21
B. Decision Rules for Assigning Group Practices to Systems	22
1. Linking group practices through the CCN approach	24
2. Linking group practices with two or more approaches.....	24
3. Name matching	25
4. Manual review	26
5. Final group practice linkage file	27
IV. Contents of the Group Practice Linkage File.....	28
A. Variables Included in the Group Practice Linkage File.....	28
B. Linking the Group Practice Linkage File to Other Data Sources	29
1. Using the TIN name to link TINs to other data sources	30
2. Using the PAC ID to link TINs to other data sources.....	30
C. Comparison of Physician Counts in the Group Practice Linkage File and 2016 Compendium.....	31
V. Caveats and Limitations.....	32
References.....	33
Appendix A. Combinations of the Four Approaches Used in the Final Set of Group Practice Linkages.....	35
Appendix B. TINs Assigned Through Manual Review	36
Appendix C. Data Dictionary	37
Appendix D. Linking the Group Practice Linkage File to the 2016 MD-PPAS Data	38
Step 1: Processing the MD-PPAS data.....	38
Step 2: Merging to the Group Practice Linkage File	39

Tables

II.1. Number of group practices (TINs) by number of physicians (NPIs)	9
II.2. Data sources used to link group practices to health systems.....	10
III.1. Summary of approaches used to link group practices to systems.....	21
III.2. Number of group practices linked to Compendium systems through each approach.....	22
III.3. Number of group practices linked to systems.....	28
IV.1. Variables included in the group practice linkage file	29
IV.2. Summary of unique and repeated TIN names	30
IV.3. Comparison of system-level physician counts	31

Figures

ES.1. Summary of decision rules for linking group practices to systems.....	4
III.1. Illustration of the CCN approach.....	14
III.2. Illustration of the hospital-based billing approach using services at HOPDs.....	17
III.3. Illustration of the Org-NPI approach	19
III.4. Illustration of the DSP approach.....	20
III.5. Decision rules for linking group practices to systems	24

Updates to This Version

The December 2020 release of the 2016 group practice linkage file contains three updates to the original version of the file released in October 2019:

1. The updated group practice linkage file includes five additional variables obtained or constructed using the 2016 Medicare Data on Provider Practice and Specialty (MD-PPAS) data. AHRQ added these variables to enable users to link the group practice linkage file to the MD-PPAS data. Once the linkage file is linked to the MD-PPAS data, users can identify the Tax Identification Numbers (TINs) and physician National Provider Identifiers (NPIs) in Compendium health systems in 2016.

The five new variables are:

- TIN name – Group practice legal name
- State – State in which the plurality of the group practice’s NPIs are located
- Number of physicians – Total number of physicians in the TIN (based on primary TIN assignment)
- Number of nurse practitioners and physician assistants – Total number of nurse practitioners and physician assistants in the TIN (based on primary TIN assignment)
- Number of line items – Total number of line items from Medicare claims billed through the TIN

The updated linkage file also includes a new flag indicating that the TIN is only found in the 2018 Medicare Advantage data, not in the 2018 fee-for-service MD-PPAS data; only TINs with a value of zero for this variable will link to the MD-PPAS data. (Refer to Section IV, Appendix C, and Appendix D for more detail.)

2. The original version of the group practice linkage file excluded 4,850 group practices with no information on possible linkages to Compendium health systems through the various data sources and methods used to conduct the linkage analysis. These group practices have 35,779 physicians in total. The updated version of the linkage file adds these group practices (none of which are linked to Compendium systems), which increases the number of group practices in the file from 36,303 to 41,153 and the number of physicians represented from 521,079 to 556,858. This update does not change any linkages to Compendium health systems.
3. When identifying the number of physicians in each group practice, we do not include physicians that switched to a new group practice during the year. In addition, in one of the approaches we use to identify group practice linkages to systems (the dominant system percentage approach, Section III.A.4), we do not include physicians switching TINs when calculating the percentage of physicians in the group practice. We identified that the original version of the group practice linkage file did not exclude all physicians switching TINs from the counts of physicians. We updated the process to identify all physicians switching group practices, and we updated the group practice linkage file. The update led to 25 fewer group practices linked to Compendium systems; the group practices have 2,570 physicians in total. Based on this change, the number of group practices linked to systems decreased from 5,355 to 5,330 (290,574 to 288,004 physicians).

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. As part of the CHSP Initiative, AHRQ and Mathematica (the initiative's Coordinating Center) developed a number of publicly available data resources for researchers, policymakers, and other stakeholders who want to understand how health systems can improve the value of healthcare.

To date, CHSP data resources include the 2016 Compendium of U.S. Health Systems (a list of 626 health systems in the United States in 2016, referred to as the Compendium in this document) and a system-hospital linkage file that links hospitals to the 626 health systems (referred to as the hospital linkage file in this document).ⁱ 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 added a physician group practice linkage file, which links group practices to their systems based on information in the data sources that indicate ownership or tight management relationships. Similar to the purpose of the hospital linkage file, the group practice linkage file contains information to enable users to link the data to the 626 health systems in the Compendium and other data sources.ⁱⁱ The group practice linkage file includes group practices both in systems and not in systems. This document describes the methodology and approach used to develop the group practice linkage file.

A. Methodology

We used the 2016 Medicare Data on Provider Practice and Specialty (MD-PPAS) data to identify the set of physician group practices (and their physicians) to be considered for linkages to the 2016 Compendium of U.S. Health Systems. The 2016 MD-PPAS data contain 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 2016.

We used 2016 Medicare Advantage data to create an analytic extract with supplemental information on NPIs and Tax Identification Numbers (TINs) billing Medicare Advantage but not Medicare Part B. In addition, 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 87,683 single-physician practices.ⁱⁱⁱ These physician group practices give us the set of group

ⁱ The Compendium identifies 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.

ⁱⁱ The technical documentation for the hospital linkage file can be found at:

https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/chsp_linkage_file_tech_doc.pdf

ⁱⁱⁱ We also excluded TINs without a TIN name in either the MD-PPAS data or Medicare Provider Enrollment, Chain, and Ownership System, which excluded 127 TINs with 303 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.

practices and their physicians to be considered for linkages to systems: 41,153 group practices with 556,858 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 follow:

1. **CMS Certification Number (CCN) approach:** We used shared TIN information in the Medicare Provider Enrollment, Chain, and Ownership System (PECOS) to link group practices to hospital CMS Certification Numbers (CCNs) and then used the hospital linkage file to link the group practices to systems. Using the PECOS data, we determined whether the group 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 indicate 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 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 relevant services occurred in a hospital-based setting, we concluded that the group practice had a strong affiliation with the hospital. Furthermore, if the hospital was owned by a health system, 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 billings for Medicare beneficiaries occurring at the HOPD settings only and a second that used such billings at all hospital-based settings, including inpatient facilities, emergency departments, and HOPDs. If the majority of a group practice's 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 the majority of physicians are *not* hospital-based specialties.

In the second version of the approach, we linked group practices with a majority of physicians in hospital-based specialties (for example, anesthesiologists and pathologists) to systems based on billings at *all* hospital-based settings (inpatient facilities, emergency departments, and HOPDs). If such a group practice reported a majority of 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 PECOS data between group practices and organizational NPIs (Org-NPIs) combined with information in the Healthcare Organization Services (HCOS) database from IQVIA to help link group practices to candidate systems. The HCOS 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.^{iv} Because the PECOS data link group practices to Org-NPIs, and the HCOS data link medical groups with Org-NPIs to systems, we were able to 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 links to an HCOS medical group through their Org-NPI and the HCOS medical group is in a system, we linked the group practice to the system through this approach. Because HCOS also links individual physician NPIs to HCOS 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 the majority 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 HCOS data with the linkages between physicians and group practices in the MD-PPAS data to help link group practices to systems. Because the MD-PPAS data links TINs to individual NPIs, and the HCOS data links individual NPIs to closely affiliated systems, we were able to merge the two data sources by individual NPI. This merge gave us the percentage of a group

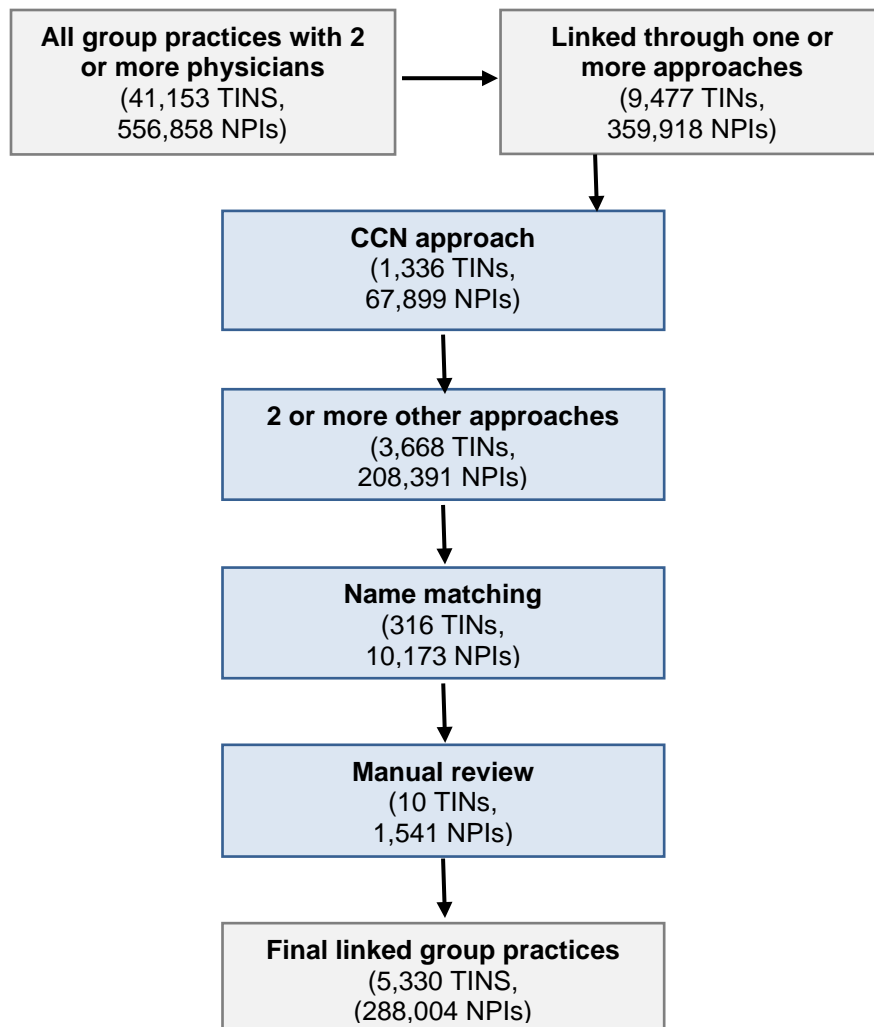
^{iv} HCOS medical groups are typically analogous to practice site locations rather than entire group practices. Thus, while there is overlap between HCOS medical groups and group practices that is useful in linking practices to systems, the two entities are not synonymous.

practice’s physicians at the TIN level (in both the MD-PPAS data and the HCOS data) linked to a system. We identified the system with the largest percentage of physicians linked to it as the dominant system percentage (DSP) and considered this system as a candidate for potential link of the given group practice to that system.

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 more than one system in the HCOS data and (2) physicians who switched to a different group practice during 2016 in the MD-PPAS data. Finally, to provide greater confidence that the group practice was linked to the correct system, we required that the majority of the group practice’s physicians (or, for larger practices, 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 41,153 group practices with 556,858 physicians in the MD-PPAS data, 9,477 group practices with 359,918 physicians 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



In the first step, we accepted group practice linkages to a system through the CCN approach, which led to accepted linkages for 1,336 group practices with 67,899 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 3,668 group practices with 208,391 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 316 group practices with 10,173 physicians.

Finally, we conducted manual review of the following group practices to identify the correct system linkages: (1) 7 large group practices (with 50 or more physicians each—1,255 physicians in total) that still linked to more than one health system after name matching and (2) group practices with any associations to eight systems that had no group practices linked through the process described above, which led to us accepting linkages for an additional 3 group practices with 286 physicians.

The final group practice linkage file includes linkages for 5,330 of the 41,153 group practices (with two or more physicians) to Compendium systems, which account for 288,004 of the 556,858 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 other data sources through two methods. First, users can link the data to the MD-PPAS data using 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 (all available in the MD-PPAS data). Alternatively, users can link the data using PECOS Associate Control ID (PAC ID) to files derived from PECOS, such as the Public Provider Enrollment data or the Physician Compare National Downloadable File.^v

^v Refer to <https://data.cms.gov/public-provider-enrollment> and <https://data.medicare.gov/data/physician-compare>.

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 then 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.

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.^{vi} 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.

To achieve AHRQ's goals, the initiative established three Centers of Excellence (CoEs) and a Coordinating Center to identify, classify, track, and compare health systems. AHRQ established CoEs at Dartmouth College, the National Bureau of Economic Research (NBER), and the RAND Corporation. Mathematica Policy Research serves as the initiative's Coordinating Center, working collaboratively with AHRQ and the CoEs to facilitate synthesis of findings on comparative health system performance, build a Compendium of U.S. Health Systems, and support dissemination of the CHSP Initiative findings broadly.

As part of the CHSP Initiative, AHRQ and Mathematica have developed a number of publicly available data resources for researchers, policymakers, and other stakeholders who want to understand how health systems can improve the value of healthcare. These data resources include a Compendium of U.S. Health Systems, a hospital linkage file, and a group practice linkage file. 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 Compendium in this document), first released in 2017, provides a list of 626 health systems in the United States in 2016 and key system attributes. The Compendium consolidates information from several data sources that indicate system ownership and provider affiliations with systems.^{vii}

The Compendium 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 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.^{viii}

The Compendium system-hospital linkage file (referred to as the hospital linkage file in this document), added in September 2018, allows users to link the 626 health systems with their member hospitals.^{ix} The hospital linkage file includes name and address information on 6,762 hospitals in systems and not in systems. In addition, the file includes hospital and system

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

^{vii} Furukawa, et al.,³ describe the methodology used to create the Compendium and use the Compendium to examine health systems in the United States in 2016.

^{viii} The technical documentation for the Compendium contains more information about how we identified qualifying health systems: https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/techdocrpt_0.pdf.

^{ix} The technical documentation for the hospital linkage file can be found at: https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/chsp_linkage_file_tech_doc.pdf.

identifiers such as the Centers for Medicare & Medicaid Services (CMS) certification number (CCN) and Compendium health system ID.

In October 2019, AHRQ added the group practice linkage file, which links group practices to Compendium health systems based on information in the data sources that indicate ownership or tight management relationships.^x The group practice linkage file includes information on 41,153 group practices (made up of 556,858 physicians) in systems and not in systems. Similar to the purpose of 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 Compendium to other data sources.

This document describes the methodology and approach used to develop the 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.

II. Data Sources

A. Identifying Group Practices and Physicians Using MD-PPAS

We used the 2016 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 2016 Compendium of U.S. Health Systems and
2. Link physician providers (identified by their National Provider Identifiers, or NPIs) to group practices (identified by Tax Identification Numbers, or TINs), which we use later to help link the group practices to systems.^{xi}

The 2016 fee-for-service MD-PPAS data contain 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 2016.^{xii} We restrict the set of providers considered in the analyses to medical doctors and doctors of osteopathy in the MD-PPAS data, which are identified by their NPIs. We refer to these providers collectively as physicians in this document.

^x 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.

^{xi} 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).

^{xii} 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>.

For each physician, the file indicates the two TINs with the largest number of line items billed by the physician in 2016; that is, the primary and secondary TINs that the physician used for billing. The data include the specialty of the physician and the number of line items billed, allowed charges, and unique patients overall and for the physicians' primary and secondary group practices. We assigned each physician to its primary TIN when developing the group practice linkage file.

In addition, we used 2016 Medicare Advantage data to create an analytic extract with supplemental information on NPIs and TINs that bill Medicare Advantage plans but not Medicare Part B. We supplemented the fee-for-service MD-PPAS data with Medicare Advantage (MA) data to ensure we included a set of group practices and physicians participating in Medicare that is as comprehensive as possible. 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).

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 Medicare Advantage data were not valid TINs.^{xiii} We also added MA-only NPIs to TINs previously identified in the fee-for-service MD-PPAS data. This step added 793 TINs that are only in the MA data and 19,618 NPIs 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 87,862 single-physician or solo practices (Table II.1).^{xiv,xv} These physician group practice linkages provided in the MD-PPAS data give us the set of group practices and their physicians to be considered for linkages to systems: 36,303 group practices with 521,709 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	128,836	100.0%	644,541	100.0%
1	87,683	68.1%	87,683	13.6%
2+	41,153	31.9%	556,858	86.4%
2-9	32,748	79.6%	118,755	21.3%
10-99	7,635	18.6%	193,067	34.7%
100-499	653	1.6%	128,352	23.0%
500+	117	0.3%	116,684	21.0%

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

^{xiii} Invalid TINs appear in the data when a valid TIN was not reported on Medicare Advantage claims.

^{xiv} We excluded TINs with a single physician to ensure the confidentiality of identifiable individuals.

^{xv} Physicians who switched from TINs during 2016 were not included in the counts of physicians for TINs.

B. Other Data Sources Used To Develop Linkages

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

Table II.2. 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	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, CCN
PECOS	Enrollment 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
HCOS	Database identifying linkages between NPIs and HCOS medical groups, NPIs and systems, and HCOS medical groups and systems	NPI, HCOS medical groups, systems	NPI, Org-NPI

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; HCOS = Healthcare Organization Services.

1. Hospital linkage file

We used information on hospital-to-system assignments in the 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 Compendium. If the hospital is linked to a health system, the file lists the name and identifier of the health system, which links to the Compendium.^{xvi}

In general, we used the 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.

^{xvi} 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/chsp_linkage_file_tech_doc.pdf.

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

We used information in PECOS 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 special permission from CMS to extract selected information from PECOS for research purposes under the CHSP project.

We used the enrollment associations (which indicate relationships between entities) in the 2016 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.

Second, PECOS data associate TINs with organizational NPIs (Org-NPIs).^{xvii} 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 98 percent of Org-NPIs link to a single TIN, and 93 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 2016 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, referred to as the hospital-based billing file in this document, 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.^{xviii}

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 setting 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 use this information to assess whether group practices and physicians billing to them should be linked to health systems.^{xix}

^{xvii} In the Healthcare Organization Services (HCOS) database, roughly half of medical groups have Org-NPIs.

^{xviii} 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 Medicare claim.

^{xix} Acumen LLC created the claims-level and TIN-CCN-level files described in this report under a contract with AHRQ.

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.

Next, we aggregated the number of DOSs to the TIN-CCN level. The claims-based 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. Thus, the file is at the TIN-hospital CCN level at this stage. 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 two rows for this TIN: one row for hospital 1 reporting its count of DOSs for the group practice and one row for hospital 2 reporting its count for the TIN.^{xx}

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 DOS services 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 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 services billed at hospital-

^{xx} The data we used to identify whether group practices occurred in a hospital-based setting include connections between each TIN and up to five hospitals where the TIN's physicians most frequently provided services, with one exception. If the fifth and sixth hospital tie, the additional hospital information is included. For example, if 50 percent of services provided by physicians in a group practice are delivered at one hospital, and 10 percent of services are delivered at each of five other hospitals, all six hospitals are included in the data, because the fifth and sixth hospitals both have values of 10 percent. If the physicians in a group practice provide services at fewer than five hospitals, information for all hospitals is included in the data. The data include complete information for all hospitals connected to the group practice's physicians for a vast majority of group practices. For the HOPD setting, 85.7 percent of TINs have complete information, and an additional 12.0 percent include between 90 and 99 percent of the TINs' billing in this setting. For all hospital-based settings, 70.8 percent of TINs have complete information, and an additional 26.2 percent include between 90 and 99 percent of the TINs' billing in these settings. For the group practices with values below 100 percent, the percentage of services provided at system hospitals is based on the top five hospitals, which is a subset of all hospitals connected to the physicians in the group practice.

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. Healthcare Organization Services (HCOS) database

Finally, we used information on medical group-to-system and physician-to-system assignments from IQVIA's Healthcare Organization Services (HCOS) database to help identify linkages between group practices and systems.^{xxi} The 2016 HCOS data include direct assignments of individual physician NPIs to medical groups and health systems (referred to in HCOS as integrated delivery networks or IDNs). HCOS 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 located 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 but are not designated as attending or IDN affiliated.

In addition, the HCOS 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. HCOS medical groups typically represent practice site locations rather than the entire physician organization. Also, the HCOS data include Org-NPIs for many of the HCOS medical groups the database identifies, which helps us link these HCOS medical groups to TINs identified in PECOS, and ultimately, the group practices to systems (described in more detail in the Methodology section below).^{xxii}

III. Methodology

We combined information from multiple approaches to link group practices to systems to create the 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.

^{xxi} IQVIA maintains two proprietary, integrated databases relevant to the study of health system performance under the umbrella of Healthcare Relational Services: OneKey Organizations, formerly known as HCOS, and OneKey Professionals, formerly known as HCPS. Throughout the document, we refer to OneKey Organizations as HCOS.

^{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 technical documentation (https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/techdocrpt_0.pdf) and Cohen, et al.⁴

A. Approaches to Linking Group Practices to Systems

1. CMS Certification Number (CCN) approach

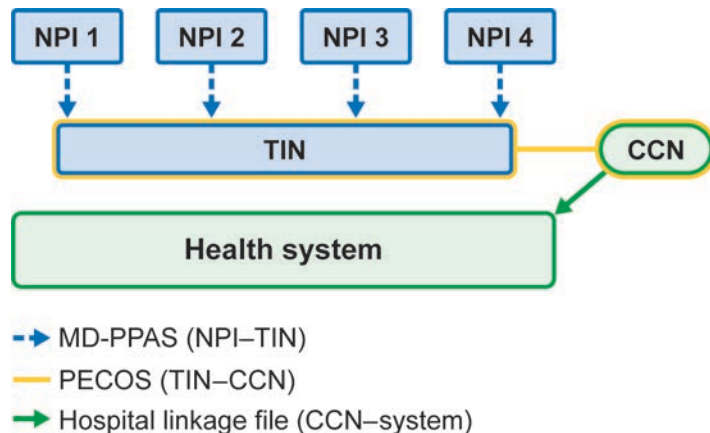
This approach used PECOS data to identify whether the group practice TIN was shared with a hospital TIN. We assumed that TINs shared by a group practice and hospital indicate common ownership by the same business entity. We used the linkages between group practice TINs and hospital TINs (and corresponding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

Finally, we excluded a small number of candidate linkages (14 group practices with 875 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 41,153 group practices in the MD-PPAS data, 2,380 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 1,336 TINs and their 67,899 NPIs through the CCN approach.

Figure III.1. Illustration of the CCN approach



2. Hospital-based billing approach

In this approach, we combined billing information in 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 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 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 +/- 7 days, and NPI of the service provider.^{xxiii}

If physicians billing under the same group practice provided a majority of their relevant services in a specific hospital-based setting, it was reasonable to conclude that the group practice had a strong affiliation with the hospital.^{xxiv} If the hospital is owned by a health system, the physician practice strongly affiliated with the hospital may be assumed to share the hospital's affiliation with the system.^{xxv}

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 the majority of physicians are in hospital-based specialties. We used services at the HOPD setting for group practices in which the majority of physicians are not in hospital-based specialties. We discuss the two versions below in more detail.

^{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} In section II.B.3, we summarize that the percentage of services provided at system hospitals is based on the top five hospitals where the TIN's physicians most frequently provided services, which for some TINs, is a subset of all hospitals connected to the physicians in the group practice (as noted in footnote xx).

Linkages based on HOPD billing. This version of the approach relies on the assumption that a group practice billing a majority of their relevant services as having occurred at a specific HOPD linked to a system (instead of an office setting owned and managed by their practice) reflects a tight relationship between the practice and the system.^{xxvi} However, basing 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 proportion 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 the majority of physicians were not in hospital-based specialties.

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

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

We identified 4,774 group practices as having a majority of physicians in hospital-based specialties (among the 12,138 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 the majority 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 2016. We applied the final restriction to require that a meaningful number of services be delivered through HOPDs, which

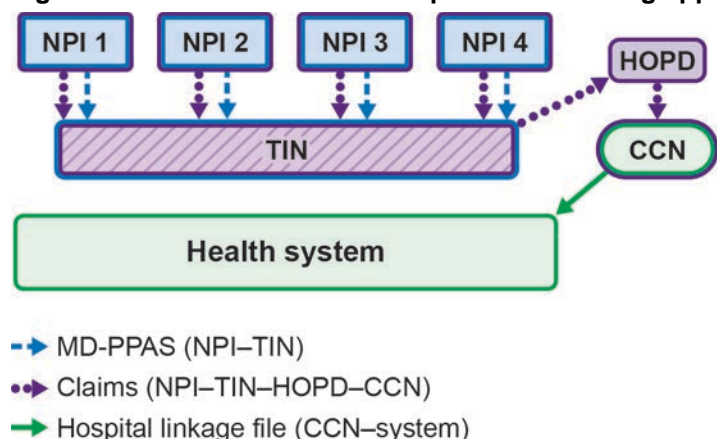
^{xxvi} When a physician service is provided in an ambulatory care setting owned by a hospital (an HOPD), the physician practice receives a greatly reduced payment that covers only the professional component of the service. However, the hospital receives an additional facility fee that results in the 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) be billed to Medicare 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

added to the confidence that the linkages to systems through HOPDs reflect 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 data with four physicians (NPIs).
- The physicians in the TIN billed a percentage of their services at HOPDs to the specific HOPD in the figure (illustrated by the dotted purple lines between the NPIs and the TIN).
- We aggregated the percentages for the NPIs 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.^{xxvii} 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

^{xxvii} 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.

in these specialties may bill a larger proportion 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 proportions 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 (not just HOPDs) to assess whether these TINs are in systems.

Hospital-based settings include HOPDs, inpatient facilities, and emergency departments. We identified candidate group practice linkages to Compendium systems when the majority of the practice's DOSs billed at hospital-based settings occurred at specific hospitals linked to a Compendium system and there were at least 10 DOSs at hospital-based settings in 2016. 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 proportion of their services in these settings all tied to the same system 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 require 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 the majority of their services but remain independent and not part of systems.

To minimize these types of inappropriate 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. We describe how we combined the information from the four approaches to make the final linkages in more detail later in this section.

We identified candidate linkages through the hospital-based billing approach for 8,332 group practices and 336,658 physicians. Of these linkages, 4,571 of the group practices and 274,539 of the physicians were linked through the HOPD setting; 3,761 of the group practices and 62,119 of the physicians were linked through all hospital-based settings.

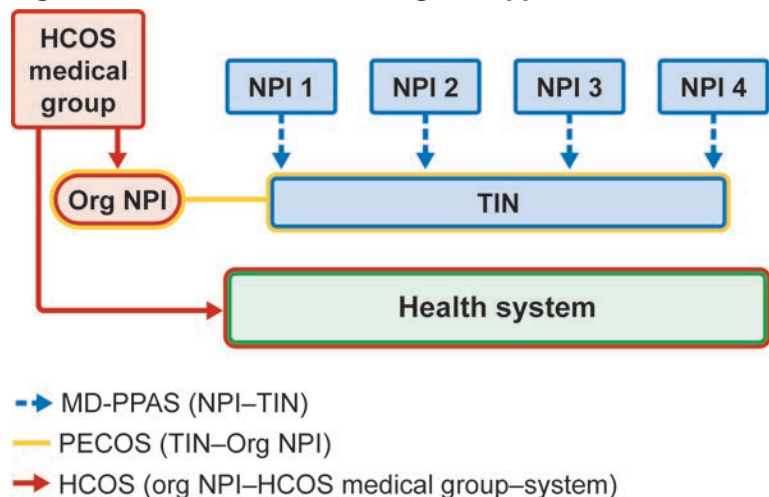
3. Organizational NPI (Org-NPI) approach

We combined the information on medical group-system assignments in the HCOS data with linkages from Org-NPIs to TINs in 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 roughly half of the HCOS medical groups reported in the HCOS data.

In Figure III.3, we show an example of a group practice (TIN) in the MD-PPAS data with four physicians (NPIs). Because the PECOS data links Org-NPIs to TINs (illustrated by the solid yellow line) and the HCOS 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.^{xxviii} We then linked TINs to candidate systems through this Org-NPI link; that is, if a TIN links to an HCOS 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.

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 comparing HCOS 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, an HCOS 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 HCOS 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 1,420 group practices and 158,146 physicians to Compendium systems through the Org-NPI approach.^{xxix}

^{xxviii} 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.

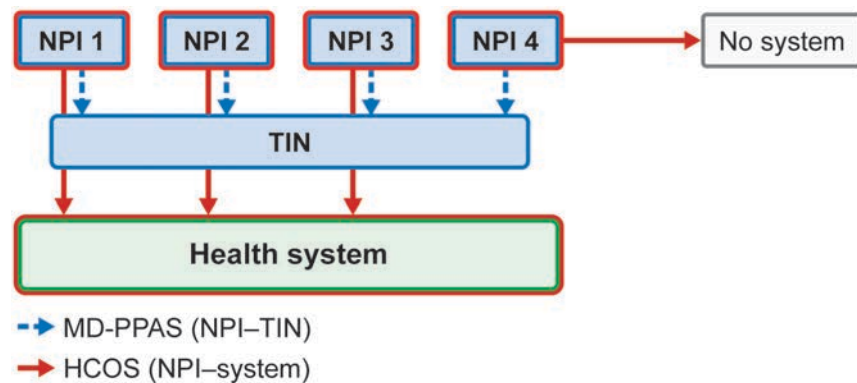
^{xxix} Before applying this restriction, we identified candidate Compendium systems for 2,092 group practices and 173,418 physicians through the Org-NPI approach.

4. Dominant system percentage (DSP) approach

We combined the NPI-level information on individual physician NPI-system assignments in the HCOS data with the linkages between physician NPIs and group practice TINs in the MD-PPAS data to help link group practices to systems. Because the MD-PPAS data link TINs to individual physician NPIs and the HCOS data link individual physician NPIs to systems, we were able to merge the two data sources by NPI. More than 90 percent of the physicians in the group practices found in MD-PPAS are also in the HCOS data (94.3 percent).^{xxx} This merge gave us the percentage of a group practice's physicians at the TIN level (in both the MD-PPAS data and the HCOS data) that are 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 HCOS 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 HCOS (to link NPIs to systems) and the MD-PPAS data (to link NPIs to TINs). The figure shows an example of a group practice (TIN) in the MD-PPAS data with four physicians (NPIs). Three of the four physicians are linked directly to the system in the diagram through the HCOS data (illustrated by the solid red arrows); the fourth physician is not assigned to a system in the HCOS 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.^{xxxi}

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 HCOS data and
2. Physicians who switched to a different TIN during 2016 in the MD-PPAS data.

^{xxx} The HCOS data included physicians not in the subset of physicians we used from the MD-PPAS data (for example, pediatricians not billing Medicare and physicians in solo practices, which we excluded from the MD-PPAS data). These physicians were not considered in this approach to linking group practices to systems.

^{xxxi} Alternatively, if NPI 4 was not in the HCOS data, the DSP would be 100 percent because all three NPIs in both data sources (NPIs 1-3) are in the system.

In the HCOS data, roughly 2 percent of the physicians (among those also in the MD-PPAS data) 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 data, 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 2016, then switching to consistently billing another TIN as the primary TIN for the remaining months of 2016. 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.

We identified 22,384 physicians as switching TINs in the MD-PPAS data. We made these exclusions so that we base 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 the majority of the TIN’s physicians (or, for larger group practices, at least 50 physicians) be linked to the candidate system through the approach. After applying this restriction, we linked 5,280 group practices and 276,668 physicians to Compendium systems through the DSP approach.^{xxxii}

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. Approximately one-quarter of group practices (9,477 or 23.0 percent) and a majority of physicians (359,918 or 64.6 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.

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); HCOS (Org-NPI-HCOS 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.

^{xxxii} Before applying this restriction, we identified candidate Compendium systems for 12,728 TINs and 424,740 physicians through the DSP approach.

Approach	Data Sources (Linkages)	Summary	Restrictions
DSP	MD-PPAS (TIN-NPI); HCOS (NPI-system)	Group practices linked to systems through their physicians	Majority of physicians or ≥50 physicians are linked through the approach.

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

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	41,153	556,858
Not linked through any approach	31,676	196,940
One or more approaches	9,477	359,918
CCN	1,336	67,899
Hospital-based billing	8,332	336,658
Org-NPI	1,420	158,146
DSP	5,280	276,668

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

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

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, both of which we have relatively high confidence in. 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 HCOS medical groups to systems through the HCOS 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:

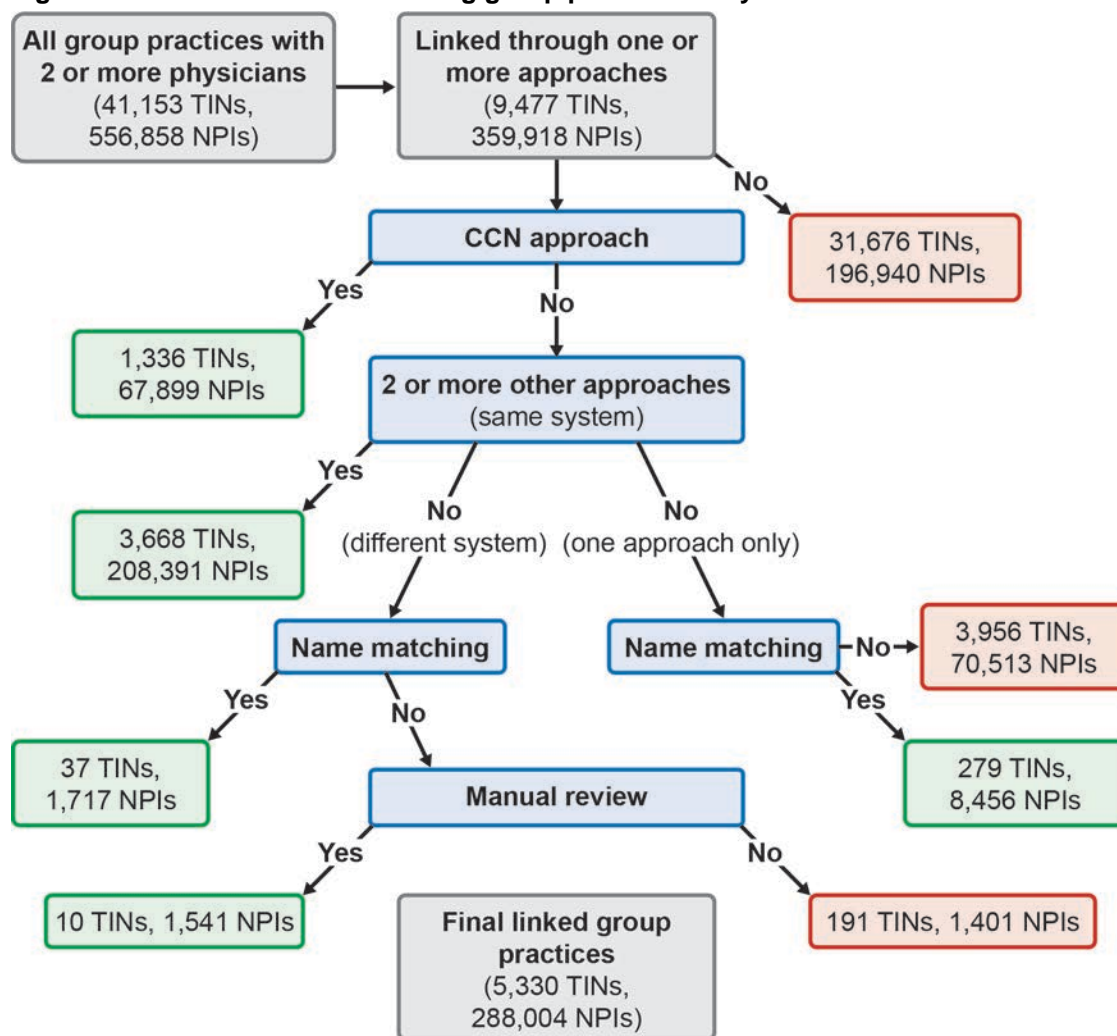
1. We accepted group practice linkages to a system through the CCN approach.
2. We accepted group practice linkages to a system through two or more of the remaining three approaches.
3. We examined whether group practice names matched system names to resolve discrepancies (group practices linked to different 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 11 large TINs (with 50 or more physicians each) that still linked to more than one health system to identify the correct system linkage. In addition, we reviewed possible linkages to eight systems that had no group practices based on the first three automated steps above. This review led to linkages for three additional group practices.

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 41,153 group practice TINs with 556,858 physician NPIs in the MD-PPAS data (restricted to TINs with two or more physicians). Of these group practices, 9,477 (359,918 physicians) linked to a system through one or more of the four approaches.

We found that 31,676 group practices with 196,940 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.

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 1,336 group practices with 67,899 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 3,668 group practices with 208,391 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 (1,820 group practices with 30,229 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 data) and health system names (from the Compendium) 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.^{xxxiii} 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 37 group practices with 1,717 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 279 group practices with 8,456 physicians through this use of name matching. Conversely, 3,956 group practices with 70,513 physicians were linked through only one approach and did not have a name match. Taken together, name matching adjudicated or confirmed linkages for 316 group practices with 10,173 physicians.

^{xxxiii} We manually reviewed a sample of the name matches to ensure that this approach generally links group practices to the correct system. All the cases reviewed linked to the correct systems.

4. Manual review

After completing the three steps to link group practices described above, we manually reviewed 8 group practices with 50 or more physicians that linked to different systems through two or more approaches but could not be resolved by name matching.^{xxxiv} In total, the 8 group practices had 1,336 physicians.^{xxxv}

Two researchers conducted web searches of the group practices and candidate systems to determine which of the systems (if any) owns or tightly manages the group practices. Specifically, they reviewed the official web pages of the group practices and systems (for example, the “about us” pages and list of practice locations), as well as other information reporting that the two were linked (for example, 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 to resolve any discrepancies.^{xxxvi}

Through the review, we linked an additional 7 group practices with 1,255 physicians to systems. Appendix B lists the 7 group practices and their systems.

Before we finalized the linkages, we compared the number of group practices and physicians in each of the 626 Compendium systems with the totals based on the linkage file. We found linkages for 618 of the 626 systems (98.7 percent), with only 8 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 four 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 an additional six group practices associated with four of the eight Compendium systems with no linked TINs.

Using the same approach in the manual review described above, two researchers conducted web searches of the group practices and candidate systems to determine if the group practices

^{xxxiv} 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.

^{xxxv} There were 195 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.

^{xxxvi} No discrepancies were found between the linkages made by the two researchers, although three cases required more discussion by the team before making a final linkage.

should be linked to the systems. We compared the linkages made by the two researchers and met to resolve any discrepancies. We determined that three of the six group practices (286 physicians) should be linked to systems, which led to three additional Compendium systems with at least one group practice; we list the group practice names and systems in Appendix B. Thus, as shown in Figure III.5, the manual review identified an additional 10 TINs with 1,541 NPIs linked to systems.

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 5,330 group practices with 288,004 physicians to Compendium systems. Thus, 13.0 percent of the 41,153 group practices (with two or more physicians) and 51.7 percent of the 556,858 physicians in these group practices are linked to Compendium systems.

Table III.3. Number of group practices linked to systems

Step in the Process	Number of Group Practices	Number of Physicians
Linkages through CCN approach	1,336	67,899
Linkages through two or more other approaches	3,668	208,391
Name matching		
To adjudicate discrepancies	37	1,717
To corroborate linkage through a single approach	279	8,456
Manual review to adjudicate additional discrepancies	7	1,255
Manual review of possible linkages to systems with no linked group practices through prior steps	3	286
Total in systems	5,330	288,004

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 final 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 MD-PPAS data, Medicare Advantage 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 MD-PPAS data that can be used to link the data to the 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, with the exception that it only includes values from MD-PPAS; that is, it does not include TIN names from the Medicare Advantage encounter data or PECOS when MD-PPAS is missing the TIN name. Finally, the file includes a flag that identifies the TINs in the group practice linkage

^{xxxvii} We obtained the TIN name from Medicare Advantage (MA) encounter data or PECOS when the name was missing from the recreated 2018 fee-for-services MD-PPAS 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 fee-for-service MD-PPAS variable. 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 MA encounter data (see section II.A); thus, these TINs have two or more physicians in the combined MA and fee-for-service MD-PPAS data files.

file found in the Medicare Advantage data but not the fee-for-service MD-PPAS data; thus, they will not link to the official MD-PPAS file. AHRQ added this set of variables to enable users to link the file to the 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.

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

Variables	Description	Source
TIN name	Group practice legal name	MD-PPAS; Medicare Advantage encounter 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 MD-PPAS data

Variables	Description	Source
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 Medicare Advantage data, not in the 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 encounter 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 these 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 then 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.

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

The vast majority of group practices in the linkage file have unique TIN names (98.5 percent) (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

	Number of Group Practices	Percentage of Group Practices
All group practices in the linkage file		
With unique names	40,529	98.5%
With names repeated one or more times	624	1.5%
Group practices in health systems		
With unique names	5,131	96.1%
With names repeated one or more times	199	3.9%
Group practices not in health systems		
With unique names	35,398	98.8%
With names repeated one or more times	425	1.2%

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 (48 of 41,153) 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.

^{xxxix} These data are available at <https://data.cms.gov/public-provider-enrollment> and <https://data.medicare.gov/data/physician-compare>.

C. Comparison of Physician Counts in the Group Practice Linkage File and 2016 Compendium

The 2016 Compendium of U.S. Health Systems contains aggregate counts of all physicians and primary care physicians. As described in detail in the Compendium’s [technical documentation](#), the counts represent the highest counts across the key data sources (HCOS, SK&A, and AHA) used to generate the 2016 Compendium. These counts varied substantially across 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 Compendium 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 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

	Number of Systems	Physicians Linked to Systems		
		Mean	Minimum	Maximum
Total physicians				
Compendium	626	742	50	20,300
Group practice linkage file	621	464	2	19,821
Difference	5	278	48	479
Primary care physicians				
Compendium	625	227	10	8,995
Group practice linkage file	607	133	1	8,266
Difference	18	94	9	729

Note: The number of systems listed is the number of systems with at least one physician (or primary care physician) reported in the Compendium or linked through a group practice in the group practice linkage file. For example, five systems do not have a group practice linked to them in the group practice linkage file. The number of systems with a primary care physician is 625 because one system is 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.

The vast majority of the 626 Compendium systems had at least one group practice assigned to them in the group practice linkage file (621 systems, or 99.2 percent). Furthermore, 531 systems (84.8 percent) would meet the inclusion criteria for qualifying as 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 Compendium physician counts 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 Compendium count was 742 compared with 464 based on the group practice linkage file. In addition, some systems had fairly large differences (for example, one system has 10,610

physicians reported in the Compendium but only 2,768 physicians based on the group practice linkage file).

While the majority of 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 813 physicians reported in the group practice linkage file but only 495 physicians based on the Compendium.

Since the total number of physicians reflected in the group practice linkage file is much lower than totals reported in the Compendium (556,858 compared with nearly 900,000 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 data, 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).

As expected, many of the systems with the largest differences in physician counts are children's systems or systems with large numbers of pediatricians, who are less likely to bill Medicare for services. Furthermore, in the Compendium data, 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 data. This difference also contributes to the higher counts in the Compendium.

Finally, the counts of primary care physicians in the Compendium could be higher because they include physicians that focus primarily on the care of hospitalized patients (hospitalists). For the Compendium, 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 data, identifies hospitalists, which enables us to remove them from counts of primary care physicians.

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 MD-PPAS data; that is, physicians billing Medicare Part B in 2016. 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 2016. For example, when comparing with the count of physicians represented in the Compendium, which includes all active physicians in 2016, the group practice linkage file reflects nearly 60 percent of physicians.

Although information on system affiliation for solo practices is not included in this file, anecdotal evidence suggests that it is unlikely that many solo practices identified in the MD-PPAS data 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 the case 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 HCOS data (NPIs and medical group locations, respectively), and 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 there was an incorrect hospital (CCN) linkage to the system. Similarly, the thresholds we use for the various approaches generally require a majority of physicians or beneficiary dates of service to link to a given system, but 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 ED 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. Also, 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 September 19, 2019.
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 September 19, 2019.
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 September 19, 2019.
4. Cohen GR, Jones DJ, Heeringa J, Barrett K, Furukawa MF, Miller D, Mutti A, Reschovsky JD, Machta R, Shortell S, Frazee 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 September 19, 2019.

5. Report to the Congress: Medicare and the Health Care Delivery System. Washington, DC: Medicare Payment Advisory Commission; June 2013. http://www.medpac.gov/docs/default-source/reports/jun13_entirereport.pdf?sfvrsn=0. Accessed September 19, 2019.

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

Approaches	Group Practice Count	Physician Count
CCN	175	1,076
CCN, DSP	121	1,184
CCN, DSP, HOPD	248	15,945
CCN, DSP, All hospital-based settings	113	1,970
CCN, HOPD	237	5,787
CCN, All hospital-based settings	70	854
CCN, Org-NPI	2	14
CCN, Org-NPI, DSP	72	1,327
CCN, Org-NPI, DSP, HOPD	242	38,336
CCN, Org-NPI, DSP, All hospital-based settings	45	1,143
CCN, Org-NPI, HOPD	8	243
CCN, Org-NPI, All hospital-based settings	3	20
DSP	34	1,726
DSP, HOPD	914	64,558
DSP, All hospital-based settings	1,820	30,229
HOPD	182	5,638
All hospital-based settings	61	1,341
Org-NPI	5	37
Org-NPI, DSP	149	12,457
Org-NPI, DSP, HOPD	729	100,429
Org-NPI, DSP, All hospital-based settings	41	731
Org-NPI, HOPD	54	2,877
Org-NPI, All hospital-based settings	5	82
Total	5,330	288,004

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

Appendix B. TINs Assigned Through Manual Review

TIN Name	Assigned Health System
Group practices originally linked to different systems	
THE SOUTHEAST PERMANENTE MEDICAL GROUP	HSI00000536/Kaiser Permanente
UNIVERSITY PHYSICIAN GROUP	HSI00001066/Tenet Healthcare Corporation
DREYER CLINIC INC	HSI00000013/Advocate Health Care
CHILDREN'S & WOMEN'S PHYSICIANS OF WESTCHESTER, LLP	HSI00000130/Boston Childrens Hospital
L&M PHYSICIAN ASSOCIATION INC	HSI00000574/Lawrence and Memorial Hospital
NORTH SHORE MEDICAL GROUP OF THE MOUNT SINAI SCHOOL OF MEDICINE	HSI00000711/Mount Sinai Health System
HARBOR MEDICAL ASSOCIATES, INC.	HSI00000813/Partners HealthCare System, Inc.
Group practices linked to systems previously with no group practices	
ARCH HEALTH PARTNERS	HSI00000809/Palomar Health
PRESBYTERIAN HEALTH PHYSICIANS	HSI00000805/PIH Health
UT REGIONAL ONE PHYSICIANS INC	HSI00000870/Regional One Health

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)
pac_id	Character	Unique PECOS Associate Control ID assigned by PECOS
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 2018 Medicare Advantage data, not in the 2018 fee-for-service MD-PPAS data; only TINs with a value of zero for this variable will link to the MD-PPAS data.

Appendix D. Linking the Group Practice Linkage File to the 2016 MD-PPAS Data

Linking the group practice linkage file with the MD-PPAS data requires (1) processing the 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 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 2016 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; counts of physicians, nurse practitioners, and physician assistants; and 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 2016 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 – this variable is 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

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;
```