2023-2024 Physician Practice Information Survey

Methodology Report

December 31, 2024

Eric Grau, Priya Bajaj, Danielle Whicher, and Martha Kovac

Submitted to:

American Medical Association
330 N. Wabash Ave., Suite 39300
Chicago, IL 60611
Attention: Carol K, Kane, Ph.D., Director, Economic and Health Policy Research
Carol.Kane@ama-assn.org

Submitted by:

Mathematica
111 East Wacker Drive, Suite 3000
Chicago, IL 60601-4303
Phone: (312) 994-1002
Fax: (609) 228-4958

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

Contents

Exhibits

Introduction

An accurate understanding of medical practices' expenses and the time physicians spend providing care to patients is critical to informing Medicare's physician payment rates. To develop estimates of practices' expenses per hour of patient care provided by physicians at the medical specialty level, the American Medical Association (AMA) contracted with Mathematica to field two large national surveys: a practice-level survey designed to gather data on practices' expenses called the Physician Practice Information (PPI) Survey and a physician-level survey called the Physician Hours Survey designed to gather data on the amount of time physician spend providing patient care.

The AMA last fielded a survey to collect data on physicians' practice expenses in 2007. Since then, there has been a shift away from solo physician practices toward large multispecialty practices (Baker et al. 2019). In addition, primary care and specialty providers have consolidated into vertically integrated health systems or hospitals (Furukawa et al. 2020; Kane 2023; Machta et al. 2020). The revised surveys and survey sampling approach described in this methodology report account for this changing physician practice landscape to develop updated estimates of practice expense per hour of patient care provided.

Before fielding the survey, the AMA contacted national medical specialty societies and state-based medical organizations to ask them to endorse the surveys. Overall, more than 170 organizations offered their support.

This document describes the approach the AMA and Mathematica took to developing the PPI and Physician Hours Surveys, selecting a national sample of practices and physicians, fielding the survey, and processing and analyzing the data.

PPI Survey Instrument Development

Overview

The purpose of the PPI Survey was to collect four main categories of information from practices:

1. General information about the practice (for example, practice location, specialties at the practice, and practice ownership)
2. The number of physicians, qualified health care professionals (QHPs), and other staff at the practice
3. The weekly hours physicians and QHPs spent providing direct patient care
4. The annual practice expenses incurred by and revenue received by physicians and QHPs related to patient care services paid for via physician payment systems

Larger practices (those with more than 10 physicians) that are not physician-owned might have difficulty estimating the amount of time their physicians spend providing direct patient care. To address that, we designed and fielded a separate Physician Hours Survey to the physicians employed by these practices instead of attempting to collect this information in the PPI Survey. We also asked physicians at small, physician-owned practices to complete the Physician Hours Survey if the PPI Survey respondent from their practice was unable to provide information on the weekly hours of direct patient care provided by the physicians. Appendix A includes a copy of the survey instrument.

Survey development

We started with a version of the survey instrument that the AMA developed and shared with Mathematica for review. In collaboration with the AMA, we revised it based on (1) the survey's intent and data requirements, (2) information learned from formative interviews the AMA conducted with large and complex organizations when planning for this survey, (3) our experience fielding practice surveys, and (4) best practices in survey design.

Revisions from the AMA's initial version of the instrument included the following:

- A screener at the beginning of the survey determined practice eligibility.
- Practices could report on their expenses and revenues for fiscal years ending on or after June 30, 2022. This meant practices could report on fiscal years ending in 2022 or 2023, depending on the most recent year for which they have complete financial data.
- Practices now had the option of combining specialties when reporting staffing, expenses, and revenue (for example, if a practice could not report on Family Medicine separately from Internal Medicine, it had the ability to combine these two and report on them together).
- Practices could report on the percentage of time physicians billed in non-facility settings, at the specialty level.
- A table created for practices that were unable to allocate expenses to QHPs. This new table asked respondents to distribute the hours of patient care provided by their QHPs to the specialties at the practice with which they work
- A question to understand the method(s) the practice used to allocate different types of expenses.
- A question to ask whether practice survey respondents would be willing to share a link to the Physician Hours Survey with the physicians at their practice (for the practices unable to provide information about weekly hours physicians spent providing direct patient care in the PPI Survey).

Instrument programming and testing

We programmed the instrument as a self-administered online survey in English using the Forsta software platform. We created a document with specifications for how to program the survey instrument and walked through these specifications with a programmer before they began work. To test the program, we used random data generator software (a software that generates a large volume of random responses to survey questions to simulate actual responses) and manually to follow all possible paths through the instrument, testing logic, reviewing data output, and confirming the instrument worked as intended. We shared log-in information and web survey links with the AMA so it could test the survey program as well.

Survey pre-test

We pre-tested the survey with a convenience sample of nine respondents to gather feedback on the survey and the respondents' experience completing it. We conducted two pre-tests with small, physician-owned practices and seven pre-tests with larger practices that were not physician-owned. Exhibit 1 provides a breakdown of pre-test practice characteristics. The AMA helped to identify and reach out to

these pre-test respondents. Mathematica sent each pre-test respondent a link to the web survey and asked them to complete the survey and then meet for a call to share feedback. During these calls, Mathematica and AMA staff probed to identify any questions or response options that caused confusion or were challenging to answer, any survey navigation issues, and any suggestions for making the survey clearer and easier to respond to.

Exhibit 1. Pre-test practice characteristics

	Total	
	Count	**%**
Ownership		
Physician-owned	6	67%
Not physician-owned	3	33%
Total	**9**	100%
Size		
Small (0–10 physicians)	2	22%
Large (> 10 physicians)	7	78%
Total	9	100%

In collaboration with the AMA, we revised the survey to improve respondents' experience and data quality. Revisions included the following:

- Minor wording adjustments and bolded text where needed helped key instructions stand out.
- We added and/or revised a number of programmed checks that appeared when respondents attempted to proceed in the survey after either entering a value outside the expected range (indicating the respondent might have misunderstood the question), or leaving half or more of the data tables with unusable data (that is, the cells were blank or had values of zero).
- We added an expense category of "billable drugs" in addition to "non-billable drugs".
- Given that the amount of time needed to complete the survey varied greatly among pre-test respondents (with estimates ranging from four to more than 40 hours), we revised our text about survey duration to say we expected the time needed to complete the survey would vary based on the size and complexity of the practice.
- We programmed a confirmation email that went to the survey respondent after they had submitted the survey.

Physician Hours Survey Instrument Development

Overview

Mathematica and the AMA developed a brief survey for physicians to collect information on the amount of time physicians spent providing direct patient care.[1] We did this because practice survey respondents

[1] We did not field the Physician Hours Survey to QHPs. We fielded this brief additional survey to physicians only.

at practices that were neither small nor physician-owned were unlikely to be able to provide this level of detailed information. Appendix A includes a copy of the survey instrument.

Survey refinement

We started with the existing survey questions for physicians that asked about hours and weeks worked that the AMA shared with Mathematica for review. In collaboration with the AMA, we revised and expanded on these questions to create the Physician Hours Survey.

Instrument programming and testing

As we did for the practice survey, we used the Forsta survey platform to program the survey in English. We created a document with detailed specifications for how to program the instrument and walked through the specifications with a programmer before they began work. We tested the program using random data generator software, as well as manually, to follow all possible paths through the instrument, testing logic, reviewing data output, and confirming the instrument works as intended. We shared log-in information and web survey links with the AMA to enable staff to test the survey program as well.

Survey pre-test

We pre-tested the instrument by asking practice survey pre-test respondents to share the hours survey link with their physicians. In total, 98 physicians from five of the pre-test practices completed the hours survey during this pre-test. We collected feedback from these physicians with a question at the end of the survey that asked respondents to share their feedback. We reviewed this feedback with the AMA and determined that no changes to the instrument were needed. However, because it took respondents only about two minutes to complete the survey, we revised our survey duration estimate from five minutes to two minutes.

Sample Frame Development and Sample Selection

Overview

To collect data on practice expense and physician hours spent providing patient care from a diverse set of practices and physicians, we created a stratified sample of practices. In the following sections, we describe how we define *practice* for the purposes of this study, as well as all aspects of the sample design, including developing the sample frame, stratification, and allocating and selecting the sample.

Throughout this section and the two sections following it, we refer to several key concepts: the sample frame, the augmented sample, the release sample, and the sample of physicians. The sample frame includes all practices identified in our source data (the Medicare Data on Provider Practice and Specialty (MD-PPAS) file and OneKey, described further below) as eligible to complete the PPI Survey. The augmented sample is a larger sample of practices than we expected to need to contact. We selected this larger sample in case our response rate to the survey was lower than anticipated and we needed to contact additional practices. The release sample refers to the practices within the augmented sample that we contacted to participate in the survey. The sample discussed in this section is a probability sample of practices, even though physicians are the unit of analysis. The sample of physicians includes all

responding physicians from the probability sample of practices, plus additional physician responses obtained from Medscape and the AMA. We obtained expenses from the practice survey and hours from the physician survey[2].

Definition of *practice* as the sampling unit

In most cases, we defined *practices* using Taxpayer Identification Numbers (TINs), entities to which physicians assigned their rights for billing and collecting payment. It was possible for a physical practice location to consist of multiple TINs, or for a TIN to be associated with multiple practice locations. We used this approach because it aligned with how the Centers for Medicare & Medicaid Services (CMS) (and many studies of physician practices) identified group practices. The TIN was the sampling unit, which was easily obtained from MD-PPAS. As such, the list of TINs available from MD-PPAS was limited to those that had physicians who had billed Medicare. MD-PPAS is available in Mathematica's Data Innovation Lab (DIL), an operational hub that supports the development and maintenance of shared data resources and infrastructure within Mathematica.

Each TIN tracks the expenses of the group's physicians. Therefore, within each TIN, we expected a practice representative to provide information on expenses for physicians within their organization. For individual-ownership practices, this representative might be the physician who owns the practice or an office administrator (Agency for Healthcare Research and Quality et al. 2020); for practices with more complex ownership, this would commonly be a chief financial officer or vice president of finance.

This framework functioned for most specialties. However, because pediatricians rarely bill Medicare due to the age of their patients, we could not use MD-PPAS to obtain data for practices that only have pediatricians. Instead, for pediatricians in practices in which no physician bills Medicare, we obtained information practice-level information from IQVIA's OneKey data set. This data set identifies practices at the site level, which might not coincide exactly with the practices as identified by TINs from MD-PPAS. For example, multiple practice sites could all have the same TIN or, conversely, a single practice might consist of several TINs. Nevertheless, we used the practice site (identified by the OneKey variable IMS_ORG_ID) as the sampling unit for pediatrician practices.

Creation of the practice sample frame

We derived the sample frame for most practices from the MD-PPAS data set and refer to it as Frame 1. To be eligible for Frame 1, TINs had to have at least one National Provider Identifier (NPI) that submitted a Medicare claim in 2020, as recorded in the MD-PPAS data set. We then used MD-PPAS in combination with other data sets, including OneKey dataset, LevinPro data, and Medicare claims data, to define other characteristics of the practices in the same frames. Specifically, we used MD-PPAS to determine the practice specialty and TIN size for practices in Frame 1; we used OneKey to determine the practice ownership type and to identify practice contact information; we used LevinPro to identify practices that had been acquired by private equity firms; and we used Medicare claims data to determine what

[2] Medscape is an online publication where physicians and other healthcare professionals come to read and access clinical content (news, condition and drug information, journal articles), CME activities, and clinical tools (pill identifier, drug interaction tool, clinical calculators). Medscape visitors are offered the opportunity to participate in market research surveys (https://www.medscape.com/).

percentage of a practice's claims are billing in a facility as this information is not available in MD-PPAS. Below is a list of the variables we created and used for explicit stratification:

- Private equity practice versus nonprivate equity practice
- Forty specialty types, consisting of a single multiple specialty category and 39 stratum specialties (as given in Exhibit 1) for single-specialty practices
- Categorized percentage of a practice's expenses in 2020 Medicare claims data that are billed in a facility (0–25%, > 25–50%, > 50–75%, > 75–100%)
- TIN size in terms of total physicians (medical doctors [MDs] or doctors of osteopathy [DOs]) (1–4, 5–10, 11–100, 101–300, 301–1,000, 1,000+)
- Ownership type (corporate-owned or physician-owned)[3]
- Multiple specialty descriptors, including:
    - More than 75 percent of physicians were primary care doctors
    - 65 to 75 percent of physicians were family medicine doctors
    - 65 to 75 percent of physicians were internal medicine doctors
    - 65 to 75 percent of physicians were orthopedic surgeons
    - 65 to 75 percent of physicians were emergency medicine doctors
    - Multiple specialty practice without one of the above descriptors

The section titled "Details of sample frame creation and sample selection" describes in detail each of these variables and how we used them to create explicit strata in [missing words? Or delete in?].

For practices that mostly had pediatricians, we created a second frame derived from the OneKey data set, referred to as Frame 2. The practice unique identifier in Frame 2 is the IMS_ORG_ID, as created by the developer of the OneKey data set, IQVIA. To be eligible for Frame 2, an IMS_ORG_ID had to consist entirely of NPIs that were not in Frame 1. From the list of variables provided above for Frame 1, none of the pediatrician practices were acquired by private equity firms and all were practices with mostly pediatricians. Facility billing levels were unavailable, so the only variables available for explicit strata were ownership type (described earlier) and IMS_ORG_ID size (with the same categories as TIN size). The section titled "Details of sample frame creation and sample selection" provides further information. The data used to construct Frame 2 are from 2023.

To create Frames 1 and 2, we implemented the following steps:

1. Create an NPI-level file using 2020 Medicare Carrier claims[4] for Frame 1 and, for Frame 2, an NPI-level file with specific pediatric specialties among NPIs that were not in Frame 1.

[3] We obtained ownership information from OneKey data. In the initial sample selection, we also considered TINs with unknown ownership for sample selection. However, given the revised eligibility criteria that excluded TINs that did not match with OneKey, these TINs were later considered ineligible.

[4] Medicare carrier claims include fee-for-service claims submitted by professional providers, including physicians, physician assistants, clinical social workers, and nurse practitioners. They are accessed through ResDAC, Research Data Assistance Center (Find the CMS Data File You Need | ResDAC).

2. Create a TIN-level file using the NPI-level file from Step 1 and MD-PPAS for Frame 1 and an IMS_ORG_ID-level file using the NPI-level file for Frame 2.
3. Assign specialty labels to Frame 1 TINs using MD-PPAS.
4. Append practice ownership information from OneKey for both frames.
5. Create the facility billing level variable using Medicare claims data for practices in Frame 1.
6. Using manual review of LevinPro acquisitions, identify which TINs are private equity TINs[5]. These matches are based on a match of practice name, plus a match of state (or neighboring state).
7. Create variables to define explicit strata for both frames.

The section titled "Details of sample frame creation and sample selection" provides details on each of these steps.

Sample selection of practices

An overview of the steps we took to select samples of practices from Frames 1 and 2 follows. For each frame, we selected an augmented sample (a sample larger than we thought we might need) for which we could determine the appropriate contact information in case the sample cases were released. From the augmented sample, we selected multiple releases, including two pilot samples, the initially planned main release, [6] and a supplemental sample; the aggregation of these four releases constitutes the entire release sample, which are the practices we attempted to contact about participating in the survey. Here are the next steps for selecting the sample:

8. Create 236 strata for sample selection of TINs in Frame 1 (a reduction from the initial 248) and five strata for Frame 2.
9. Assign the strata to NPIs in the NPI-level files described in Step 1 for both frames.
10. To prepare to select the augmented sample for both frames, allocate projected completed practice-level surveys and number of augmented sample cases to strata assuming a 10 percent completion rate.
11. Using equal probability sampling within the explicit strata, select an augmented sample of TINs.
12. Identify a contact person for each TIN.
13. Obtain a list of physicians for all TINs in the augmented sample.
14. To prepare to select the release sample, create substrata within the augmented sample for TINs in Frame 1 known to have an email address and allocate anticipated completed interviews for release sample cases within both the original strata from Step 8 and the substrata.
15. Select a release sample of TINs from augmented sample of Step 11.

[5] LevinPro maintains a database of health care acquisitions data that includes information on private equity ownership (LevinPro HC - Healthcare M&A Intelligence - Levin Associates).

[6] In the initially planned main release, some practices were later deemed ineligible prior to data collection. We selected a replacement sample to replace these ineligible cases. This sampling step is described in Step #19 of sample selection.

16. Determine which pre-test cases were in the Frame 1 sample frame, which were in the augmented sample, and which were in the released sample.
17. Exclude TINs in the Frame 1 sample that did not have a match with the OneKey file.
18. Separate the release sample in Frame 1 into pilot samples and the remaining sample.
19. Reselect sample cases for the release sample in Frame 1 to replace ineligible cases.
20. Select a supplemental sample from Frame 1.
21. Incorporate volunteer practices into the Frame 1 sample.
22. Add relevant Clinician Practice Information (CPI) frame cases to the intended population.

The section titled "Details of sample frame creation and sample selection" provides details about Steps 8 through 22. Step 8 includes an eligibility criterion added after we completed the initial sample frame creation and sample selection: we required the TINs in Frame 1 to have a match with the OneKey data set because we surmised that either (1) the practices had closed between 2020 and 2023 and would be ineligible anyway or (2) their lack of contact information from OneKey would have made the open practices extremely difficult to contact. This reduced the number of eligible TINs in Frame 1 and it reduced the size of the augmented sample selected from that frame.

Sample of physicians

Our plan was to obtain physician sample cases from the sampled practices. Because of the low response rate to the PPI Survey, we supplemented this with physician survey responses from Medscape (physicians from our sampled practices who happened to be in the Medscape panel) and from the AMA. The survey administration section describes the process of bringing these responses together.

Details of Sample Frame Creation and Sample Selection

Sample frame creation

This section provides details on the seven steps for creating the sample frame that we outlined in the introduction. **Bold text** identifies each step and is followed by the description of that step.

Create an NPI-level file using 2020 Medicare Carrier claims for Frame 1 and, for Frame 2, an NPI-level file with specific pediatric specialties among NPIs that were not in Frame 1 (Step 1)

Medicare defines 64 distinct medical specialties. When we examined the number of practices within each of these 64 specialty areas to inform the sampling process, we determined some specialties had very small counts and therefore it was not practical to allocate sample cases to practices based on all these specialties. As a result, the AMA collapsed the 64 specialties into 47 specialties of interest for analysis. Even after collapsing from 64 to 47, some of these specialties were still very small. As a result, the AMA further collapsed the 47 specialties into 39 specialties for explicit stratification in sampling (Exhibit 1). When the AMA finalized this mapping, the first step for creating the sampling frame of practices involved accessing 2020 Carrier claims data from Medicare, rolling up the claims data to the NPI level, and assigning specialties to NPIs based on the two groupings of specialty definitions given in Exhibit 2. Hereafter, we will refer to these 47 specialties as practice sampling specialties and the 39 specialties as

stratum specialties.[7] We used the 2020 version of the claims data to coincide with the latest version of MD-PPAS (MDPPAS_V24_2020) that was available at the time we created the sampling frame. Both the 2020 Carrier claims data and the 2020 version of MD-PPAS (MDPPAS_V24_2020) were available through DIL's access to the Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC). We limited the file to NPIs billing under TINs in the 50 United States plus the District of Columbia.

Exhibit 2. Crosswalk of 64 specialties to 47 specialties for analysis and 39 specialties for sampling

CMS specialty code	Code description	Grouped 47 practice sampling specialties	Grouped 39 stratum specialties
03	Allergy/Immunology	Allergy & Immunology	Allergy & Immunology
05	Anesthesiology	Anesthesiology	Anesthesiology
21	Cardiac Electrophysiology	Cardiac Electrophysiology	Cardiology
06	Cardiology	Cardiology	Cardiology
C7	Advanced Heart Failure and Transplant Cardiology	Cardiology	Cardiology
D8	Adult Congenital Heart Disease	Cardiology	Cardiology
33	Thoracic Surgery	Cardiothoracic Surgery	Cardiothoracic Surgery
78	Cardiac Surgery	Cardiothoracic Surgery	Cardiothoracic Surgery
07	Dermatology	Dermatology	Dermatology
D7	Micrographic Dermatologic Surgery	Dermatology	Dermatology
93	Emergency Medicine	Emergency Medicine	Emergency Medicine
C8	Medical Toxicology	Emergency Medicine	Emergency Medicine
46	Endocrinology	Endocrinology	Endocrinology
08	Family Medicine	Family Medicine	Family Medicine
23	Sports Medicine	Sports Medicine	Family Medicine
10	Gastroenterology	Gastroenterology	Gastroenterology
01	General Practice	General Practice	General Practice
28	Colorectal Surgery	Colorectal Surgery	General Surgery
02	General Surgery	General Surgery	General Surgery
91	Surgical Oncology	General Surgery	General Surgery
98	Gynecological/Oncology	General Surgery	General Surgery
38	Geriatric Medicine	Geriatric Medicine	Geriatric Medicine
40	Hand Surgery	Hand Surgery	Hand Surgery
82	Hematology	Hematology/Oncology	Hematology/Oncology
83	Hematology/Oncology	Hematology/Oncology	Hematology/Oncology
90	Medical Oncology	Hematology/Oncology	Hematology/Oncology
C9	Hematopoietic Cell Transplantation and Cellular Therapy	Hematology/Oncology	Hematology/Oncology
17	Hospice and Palliative Care	Hospice & Palliative Medicine	Hospital Based
C6	Hospitalist	Hospitalist	Hospital Based

[7] Initially, the AMA and Mathematica intended to create explicit strata for practices based on the 47 practice sampling specialties. However, due to the small stratum sizes for some of the 47 specialties, we decided to define explicit strata based on 39 collapsed specialties, with the intention of using information about the original 47 specialties in weighting. With the low response rates, we found it necessary to collapse the 39 specialties even more, as shown in Exhibit 8 in the weighting section.

CMS specialty code	Code description	Grouped 47 practice sampling specialties	Grouped 39 stratum specialties
44	Infectious Disease	Infectious Disease	Hospital Based
81	Critical Care (Intensivist)	Critical Care Medicine	Hospital Based
11	Internal Medicine	Internal Medicine	Internal Medicine
84	Preventive Medicine	Internal Medicine	Internal Medicine
12	Osteopathic Manipulative Medicine	Osteopathic Manipulative Medicine	Internal Medicine
C3	Interventional Cardiology	Interventional Cardiology	Interventional Cardiology
09	Interventional Pain Management	Interventional Pain Medicine	Interventional Pain Medicine
94	Interventional Radiology	Interventional Radiology	Interventional Radiology
39	Nephrology	Nephrology	Nephrology
13	Neurology	Neurology	Neurology
14	Neurosurgery	Neurosurgery	Neurosurgery
16	Obstetrics/Gynecology	Obstetrics/Gynecology	Obstetrics/Gynecology
18	Ophthalmology	Ophthalmology	Ophthalmology
20	Orthopedic Surgery	Orthopedic Surgery	Orthopedic Surgery
04	Otolaryngology	Otolaryngology	Otolaryngology
85	Maxillofacial Surgery	Otolaryngology	Otolaryngology
72	Pain Management	Pain Medicine	Pain Medicine
22	Pathology	Pathology	Pathology
D3	Medical Genetics and Genomics	Pathology	Pathology
37	Pediatric Medicine	Pediatrics	Pediatrics
25	Physical Medicine and Rehabilitation	Physical Medicine & Rehabilitation	Physical Medicine & Rehabilitation
24	Plastic and Reconstructive Surgery	Plastic Surgery	Plastic Surgery
26	Psychiatry	Psychiatry	Psychiatry
27	Geriatric Psychiatry	Psychiatry	Psychiatry
79	Addiction Medicine	Psychiatry	Psychiatry
86	Neuropsychiatry	Psychiatry	Psychiatry
29	Pulmonary Disease	Pulmonary Disease	Pulmonary Disease
92	Radiation Oncology	Radiation Oncology	Radiation Oncology
36	Nuclear Medicine	Nuclear Medicine	Radiology
30	Diagnostic Radiology	Radiology	Radiology
66	Rheumatology	Rheumatology	Rheumatology
C0	Sleep Medicine	Sleep Medicine	Sleep Medicine
34	Urology	Urology	Urology
76	Peripheral Vascular Disease	Vascular Surgery	Vascular Surgery
77	Vascular Surgery	Vascular Surgery	Vascular Surgery

We mapped the specialty codes available in the MD-PPAS variable primary specialty of NPIs (spec_prim_1) to one of the 47 physician sampling specialties. We first restricted the universe of NPIs to those with a 'spec_broad' code of 1 through 6, which are the following broad categories: (1) primary care, (2) medical specialty, (3) surgical specialty, (4) obstetrics/gynecology with no primary care specialty, (5) hospital-based specialty (includes designated hospitalists), and (6) psychiatry. Then we dropped cases in which

'spec_prim_1' was equal to D4 (Undersea and Hyperbaric Medicine) or 99 (Other) because they did not have unique matches to the mapping document. The sample retained only physicians registered in the 50 states and the District of Columbia. We added an additional criterion of eligibility after we created the initial frame, which excluded NPIs from MD-PPAS that did not have at least one affiliation with a qualifying organization in OneKey. Initially, the file included 670,563 NPIs, but after adding the eligibility criterion the number of NPIs declined to 664,695 eligible NPIs.

For Frame 2, we created a file of NPIs from OneKey that did not match with MD-PPAS, in which the practice's state (the state of the IMS_ORG_ID) was within the 50 United States or the District of Columbia. We then obtained the primary specialty for each NPI and only retained NPIs in the OneKey universe that matched the OneKey pediatric specialties in Exhibit 3.

Exhibit 3. OneKey pediatric specialties used for Frame 2

OneKey specialty code	Code description
PSM	Pediatric Sports Medicine
ADL	Adolescent Medicine
CAP	Child Abuse Pediatrics
DBP	Developmental-Behavioral Pediatrics
MPD	Internal Medicine/Pediatrics
NPM	Neonatal-Perinatal Medicine
PD	Pediatrics

For Frame 2, 28,807 NPIs met these criteria.

Create a TIN-level file using the NPI-level file from Step 1 and MD-PPAS for Frame 1, and an IMS_ORG_ID-level file using the NPI-level file for Frame 2 (Step 2)

We gathered all TINs from the file in Step 1 and linked them to TINs in MD-PPAS. This was the practice-level TIN frame for Frame 1. MD-PPAS defines TIN1—a "pseudoTIN"—for physicians based on the majority of their charges. We retained only NPIs with an associated pseudoTIN derived from the TIN1 field. We then assigned an internal random string called RANDOM on the CCW to allow for pseudoTIN-level extracts. We used RANDOM as our identifier for each practice as defined by the TIN1 field. Finally, we excluded all TINs for which their constituent NPIs were not found in the OneKey data set. The final TIN frame included 106,618 TINs. Because a physician could be affiliated with multiple TINs, the list of TIN–NPI combinations was larger than the list of NPIs given in Step 1. In total, the frame included 853,517 TIN–NPI combinations (664,695 unique NPIs).[8]

Frame 2 comprised only those IMS_ORG_IDs in which 75 percent or more of the NPIs were pediatricians according to OneKey specialty tables. This resulted in a frame of 14,771 IMS_ORG_IDs, with 37,302 IMS_ORG_ID-NPI combinations and 28,807 unique NPIs. Therefore, the total number of practices (TINs for Frame 1, and IMS_ORG_IDs for Frame 2) is 121,389, with 693,502 unique NPIs.

[8] Before applying the final criterion (excluding TINs with NPIs not found in OneKey) there were 113,160 TINs with 860,170 TIN–NPI combinations and 670,563 unique NPIs.

Assign specialty labels to TINs using MD-PPAS (Step 3)

We are interested in the estimates of practice expenses and physician patient care hours for specialists corresponding to one of the 64 Medicare specialties. Because we sampled practices instead of physicians, we needed a way to identify the distribution of the specialties of the physicians within the practices. Even though we sampled practices, knowing the distribution of specialties enabled us to allocate the sample of practices so we could anticipate how many physicians within each specialty would be in the sample. We assigned specialty labels to TINs based upon the preponderance of NPI specialties in the TIN. For the purposes of this study, using our sample frame data for Frame 1, we defined *single-specialty practices* as those in which at least 75 percent of their physicians identified as members of a single specialty. We categorized the remaining practices as multispecialty practices. Because Frame 2 is composed only of IMS_ORG_IDs in which 75 percent or more of the NPIs were pediatricians, all practices in Frame 2 were single-specialty practices.

Append ownership information from OneKey for both frames (Step 4)

The only information the MD-PPAS provides about a TIN's location is the state, and it provides no information about practice ownership. Because of this, we needed to obtain ownership information from OneKey. There is no common merging variable between OneKey and the file generated using Carrier claims and MD-PPAS at the TIN level. Therefore, we completed this merge using NPI—at the physician level. We linked TINs from the TIN frame to IMS_ORG_IDs from OneKey by determining which IMS_ORG_ID shared the greatest number of NPIs with each TIN.

Next, we used ownership categories from the OneKey business table to assigned organizations at the IMS_ORG_ID level and associated pseudoTIN to one of three ownership categories. The following ownership fields were available from OneKey:

- corp_parent_ims_org_id: identifier used to associate OneKey level organizations with parent corporations
- primary_cot_id: this was the numeric code for primary class of trade

If the primary_cot_id value indicated "Corporate Parent – Integrated Delivery Network," then we assigned the associated IMS_ORG_ID and mapped pseudoTIN to an ownership category of "Parent: Integrated Delivery Network." If this value did not indicate "Corporate Parent – Integrated Delivery Network" but there was an associated parent corporation identifier, then the ownership category was "Parent: Other Corporate." If there was no associated parent corporation identifier, then the ownership category was "Independent: physician-owned."[9]

Create the facility billing variable using claims data (Step 5)

The percentage of claims billed in a facility is calculated as the total amount billed in a facility setting for a TIN, divided by the total amount billed, multiplied by 100. The 2020 Medicare claims data were available through the DIL's access to the CCW VRDC. On the VRDC, these claims-level data are accessed as

[9] Source: OneKey Reference Data Integrated Prescriber and Organization Reference Data OneKey Relational (ICADM) Data Dictionary Version 2.14.

Research Identifiable File (RIF) Part B line-level carrier files. The line-item claims and associated charge amounts (line_alowd_chrg_amt) were determined to be in a facility setting if the LINE_PLACE_OF_SERVC_CD value was equal to one of 21 possible codes given in Exhibit 4. Appendix B describes each of the codes in detail.

Exhibit 4. Facility-based place-of-service codes with brief descriptions

LINE_PLACE_OF_SERVC_CD value	Description
02	Telehealth
05	Indian Health Service - Free-standing Facility
06	Indian Health Service - Provider-based Facility
07	Tribal 638 - Free-standing Facility
08	Tribal 638 - Provider-based Facility
09	Prison/Correctional Facility
21	Inpatient Hospital
22	Outpatient Hospital
23	Emergency Room - Hospital
24	Ambulatory Surgical Center
26	Military Treatment Facility
31	Skilled Nursing Facility
34	Hospice
41	Ambulance - Land
42	Ambulance - Air or Water
51	Inpatient Psych Facility
52	Psychiatric Facility - Partial Hospitalization
53	Community Mental Health Center
56	Psychiatric Residential Treatment Center
58	Unassigned
61	Comprehensive Inpatient Rehabilitation Facility

Source: Chronic Condition Data Warehouse CODEBOOK Medicare Fee-For-Service Claims (for Version K) April 2020 Version 1.6.

Charges allowable for the percentage of facility spending field were also restricted to the following:

1. Exclude type of service 'F' (Ambulatory Surgery Center [ASC] facility fees) and
2. Make the following restrictions based on linking codes in the Healthcare Common Procedure Coding System (HCPCS) to the national Relative Value Unit (RVU) file available here: https://www.cms.gov/medicaremedicare-fee-service-paymentphysicianfeeschedpfs-relative-value-files/rvu20d
   - Medicare Physician Fee Schedule (MPFS) status code of 'A', 'J', or 'T' (active, anesthesia or injections) or
   - MPFS status code of 'C' or 'R' (carrier-priced or restricted) if the HCPCS is a Current Procedural Terminology (CPT) code (numeric HCPCS) or 'G-code' (first character of HCPCS is 'G').

Identify private equity TINs by manual review of LevinPro acquisitions (Step 6)

Mathematica received a one-year subscription to LevinPro that contained a database of private equity acquisitions. The data source contained a universe of private equity acquisitions from 1995 through March 2023 (as of the date of review in March 2023). When filtered to distinct acquisition target names and locations, 375 observations remained. Both the MD-PPAS universe and this LevinPro data went through standard string cleaning before attempting to match the target entity name to pseudoTIN name.

For those pseudoTINs that matched to acquisition targets, we manually compared the Core Based Statistical Area of the majority of pseudoTIN practitioners to the target entity location. If the state matched or was a geographic neighbor, we determined the pseudoTIN to be a private equity acquisition. In the main sample, we determined 36 pseudoTINs to be private equity acquisitions.

Create variables for explicit stratification (Step 7)

All the variables needed to create explicit strata are available from the steps described before. Since there were very few practices that we identified as private equity owned in Frame 1, we first created a separate stratum for these practices. From there we created strata using the remaining variables, prioritizing specialty and facility billing as these were the most important for the analysis.[10] As such, these variables garnered the highest priority for explicit stratification in Frame 1. In addition, we incorporated TIN size, ownership category, and several descriptors for some multiple specialty categories. The most granular categorizations used for each variable are as follows:

1. Private equity practice versus nonprivate equity practice
2. Forty specialty types, including a single multiple specialty category and 39 stratum specialties (as given in Exhibit 1) for single-specialty practices
3. Categorized percentage of a practice's expenses in 2020 Medicare claims data that are billed in a facility (0–25%, > 25–50%, > 50–75%, > 75–100%)
4. TIN size in terms of total physicians (MDs or DOs) (1–4, 5–10, 11–100, 101–300, 301–1,000, 1,000+)
5. Ownership type (corporate-owned, physician-owned, unknown)[11]
6. Multiple specialty descriptors, including:
    a. More than 75 percent of physicians were primary care doctors
    b. 65 to 75 percent of physicians were family medicine doctors
    c. 65 to 75 percent of physicians were internal medicine doctors
    d. 65 to 75 percent of physicians were orthopedic surgeons

[10] At the beginning of the project, there was considerable interest in the impact of private equity ownership on expenses per hour. However, the number of TINs that were among the private equity acquisitions was so small that AMA eventually determined we could not use this information. We created a separate stratum for private equity practices across all other explicit stratification variables, but with no responses in the stratum, we could not use this information.

[11] All TINs in the unknown category were ineligible; if they had unknown ownership, they did not match with the OneKey data.

e. 65 to 75 percent of physicians were emergency medicine doctors

f. Multiple specialty practice without one of the above descriptors

The granular categorizations of the variables available in Frame 2 are the following:

1. Private equity practice versus nonprivate equity practice (no private equity practices in Frame 2)
2. Pediatric single-specialty practices (only one specialty in Frame 2)
3. IMS_ORG_ID size in terms of total pediatricians (1–4, 5–10, 11–100)
4. Ownership type (corporate-owned, physician-owned)

Select sample of practices

This section describes the selection of the sample of practices, defined as either TINs for Frame 1 or IMS_ORG_IDs for Frame 2. Initially, we selected more practices than we expected to need to account for expected differential response and eligibility rates in the sampling strata. We separated this large sample, called an augmented sample, into two parts: (1) a sample that we definitely intended to release; and (2) a reserve sample, with a random number affixed to each reserve case that we used to identify supplemental cases for release in case the initial release was not large enough. The final release sample included both the initial release sample and a single supplemental sample selected from the reserve cases. Steps 8 through 22 describe the selection of this sample.

Create 236 strata for sample selection of TINs in Frame 1 (a reduction from the initial 248) and 5 strata for Frame 2 (Step 8)

Each stratum required collapsing levels with at least some of the variables above so that each stratum had at least 20 TINs in the frame. We created 236 strata for Frame 1 (initially, there were 248, but applying the OneKey eligibility criterion reduced the number) and five strata for Frame 2.[12] For the strata in Frame 1, only one stratum involved all 36 private equity TINs, 191 strata involved single-specialty TINs, and 44 strata involved multiple-specialty TINs. About half of the strata were defined by TIN sizes of 10 or fewer physicians (22 of the 44 multiple-specialty strata and 96 of the 191 single-specialty strata).

We defined the five strata in Frame 2 in terms of the two variables that were available: ownership type and size of practice, as defined by IMS_ORG_ID. The vast majority of practices in Frame 2 were very small (about 85 percent had four or fewer physicians; more than 98 percent had 10 or fewer).

Appendix C provides detailed lists of the 253 strata for both Frame 1 and Frame 2, with the counts of practices in the population within each stratum.[13]

[12] These steps are laid out chronologically, and we initially created 248 strata for Frame 1. We did not reduce the number of strata from 248 to 236 until Step 18. The reduction occurred in Step 18 after applying the eligibility criterion requiring that NPIs have a match with OneKey data, 12 strata were removed as all TINs within those strata were ineligible.

[13] Only 241 strata had positive counts. The 12 removed strata had 0 counts in Appendix C.

Assign the strata to NPIs in the NPI-level files in Step 1 for both frames (Step 9)

Even though the practice (TIN for Frame 1, IMS_ORG_ID for Frame 2) is the sampling unit, there is analytic interest at both the practice and physician levels. As such, it was important to have a count of the total number of physicians within each stratum in addition to the number of practices. Appendix D provides counts of practices by explicit strata in each frame, along with the counts of the allocated number of completed interviews (Step 10).

To prepare for the selection of the augmented sample for both frames, allocate projected completed practice-level surveys and number of augmented sample cases to strata assuming a 10 percent completion rate (Step 10)

We selected an augmented sample and selected replicates from that augmented sample for the projected total number of released cases. We intended the augmented sample to be large enough that we would not release all cases, no matter how low the response rate might eventually be.

We planned the sample design with the intention of obtaining more than 3,200 completed surveys in Frame 1 and 100 completed surveys in Frame 2. When allocating the planned completed surveys to strata, we wanted to ensure enough completed surveys for subgroups of interest, while minimizing unequal weighting across strata, to reduce the design effect due to unequal weighting. Therefore, we tried to hew as closely as possible to a proportional allocation, though some of the smaller strata had an allocation that was necessarily higher than what we would have obtained in a proportional allocation. We obtained the number of augmented sample cases (initially 30,140 TINs in Frame 1, 1,000 IMS_ORG_IDs in Frame 2) by dividing the planned number of completed surveys by 0.10, which reflects an assumed worst-case scenario of an approximate 10 percent completion rate.[14]

Using equal probability sampling within the explicit strata, select an augmented sample of TINs (Step 11)

We created each stratum based on collapsed versions of the variables listed in Step 7; for some specialties, we collapsed variables so they did not contribute to the explicit stratification at all. For Frame 1, we sorted TINs within each stratum according to the following variables, which in many cases were more granular forms of the variables used for explicit stratification:

1. 47 practice specialties
2. 4 facility setting categories (0–25, > 25–50, > 50–75, > 75–100)
3. TIN size categories not used in explicit stratification due to collapsing
4. Ownership categories not used in explicit stratification due to collapsing, plus the corporate category is split into two types of corporate ownership: vertically integrated versus not vertically integrated

[14] These steps are laid out chronologically, and we initially selected 30,133 TINs for Frame 1, but it was reduced to 28,529 after removing TINs in Step 18 after applying the eligibility criterion requiring that NPIs have a match with OneKey data. (The numbers initially indicated as selected were 30,143 and 28,539, but these numbers include 10 pre-test cases.) This occurred both by removing the TINS within the 12 ineligible strata and by removing ineligible TINs within the 236 eligible strata. As will be clear in the section on weighting, our worst-case scenario of a 10 percent completion rate was not the worst case.

5. TIN size (not categorized)

We selected, with equal probability, the augmented sample of 28,536 TINs.[15] We sorted the variables listed above within explicit strata ensuring that the distribution of TINs for these variables mirrored that of the population as much as possible. This process, called implicit stratification, is most effective with the sorting variables highest in the list given above.

The process for Frame 2 was similar, but the list of implicit stratification variables available for sorting within explicit strata was more limited:

1. IMS_ORG_ID size categories not used in explicit stratification due to collapsing
2. Ownership categories not used in explicit stratification due to collapsing, plus the corporate category is split into two types of corporate ownership: vertically integrated versus not vertically integrated
3. IMS_ORG_ID size (not categorized)

We selected an augmented sample of 1,000 IMS_ORG_IDs, which we sorted by the three variables listed above within explicit strata.

Identify a contact person for each TIN (Frame 1) or IMS_ORG_ID (Frame 2) (Step 12)

We needed to identify the person or people affiliated with a OneKey IMS_ORG_ID that linked to each TIN in the augmented sample from Frame 1 who was or were most likely able to answer to survey. To do this, we used the OneKey "contact_affiliation_fact" table. We linked the contact title code (title_group_code) in that table with the contact description in the "professional_title" table, using the variable "title_group_desc," and identified a prioritized list of contact types associated with one or more of the IMS_ORG_IDs linked to that TIN:

1. Head of finance
2. Head of administration (chief executive officer or president)
3. Head of administration (administrator, general manager, or other)
4. Head of operations
5. Office manager

If there was more than one person with any of these titles, we kept them all. If it was not possible to identify an individual with any of these titles, we looked for someone with any of the following titles, in priority order:

1. Vice president
2. Business office director
3. Chief of staff
4. Medical director
5. Revenue cycle manager

[15] See the footnote in Step 10 regarding the initial selection of 30,140 TINs for Frame 1.

If it was not possible to identify any of the preceding types of professionals, we identified all staff affiliated with the IMS_ORG_IDs linked to each TIN. For the individuals identified, we obtained an email address (if available) and the mailing address for the affiliated IMS_ORG_ID.

For Frame 1, we implemented this process for all IMS_ORG_IDs affiliated with TINs in the augmented sample; for Frame 2, we implemented it only for the IMS_ORG_IDs in the release sample. For TINs with only one affiliated NPI or that had a provider's name listed at the TIN name, we identified the physician contact information if the professional contact was not available.

Obtain a list of physicians for all TINs in the augmented sample in Frame 1, and IMS_ORG_IDs in Frame 2 (Step 13)

We used the MD-PPAS data to identify all physicians that billed to a TIN in the augmented sample in Frame 1 and we used the OneKey data to identify all physicians affiliated with an IMS_ORG_ID in augmented sample for Frame 2. About 48 percent of the NPIs in the augmented sample from Frame 1 had at least one usable email address, a small proportion of which had multiple email addresses, and 45 percent of the NPIs in the augmented sample from Frame 2 had a usable email address. Even though we obtained the NPIs for Frame 1 from MD-PPAS and from OneKey for Frame 2, we obtained the email addresses for the NPIs from OneKey for both frames.

To prepare for the selection of the initial release sample, create substrata within the augmented sample for TINs in Frame 1 that are known to have an email address and allocate anticipated completed interviews and number of release sample cases within both the original strata from step 8 and the substrata (Step 14)

Within the augmented sample, we substratified the sampled cases in Frame 1 by whether we could identify someone from Step 12 who had an email address we could use to contact them to complete the survey. We did not employ any substrata for Frame 2.

The allocation of anticipated completed interviews assumed we would obtain about 3,200 completed interviews in Frame 1 and 100 in Frame 2. We allocated the 3,300 so that the probability of selecting a practice with an email address was larger than the probability of selecting a practice without one. We obtained the number of released sample cases (initially 10,709 TINs in Frame 1 not including pre-test cases, and 336 IMS_ORG_IDs in Frame 2) by dividing the planned number of completed surveys by 0.30, assuming a 30 percent completion rate. Appendix D provides the number allocated to each substratum.

Select an initial release sample of TINs from augmented sample of Step 11 (Step 15)

We selected an equal probability sample of 10,221 TINs from the selected augmented sample cases within the 236 strata in Frame 1 and 336 IMS_ORG_IDs in Frame 2.[16]

[16] These steps were chronological; the initial number of TINs selected was 10,709 from the initial 248 strata, but 488 TINs were later removed due to the eligibility criterion that required matching the TIN's NPIs to the OneKey data set. (The number initially indicated as selected was 10,719, but this number included 10 pre-test cases—see Step 17.) We selected 226 replacement TINs in Step 20 from the 28,539 – 10,221 = 18,308 TINs that remained in the augmented sample.

Determine which pre-test cases were in the Frame 1 sample frame, which were in the augmented sample, and which were in the released sample (Step 16)

The AMA selected 10 practices with which it had a previous relationship to fill out the survey as pre-test practices. Because we did not select these practices from either sample frame (Frame 1 or Frame 2), we had to determine how the practices matched up with the frames. None of the practices were pediatric practices, excluding any possible matches with Frame 2. However, there was some ambiguity in this process of matching with Frame 1, as it was not possible to determine whether the pre-test practices corresponded directly to the TINs we identified as sampling units. Nevertheless, we determined that one of the 10 practices was part of the initial release sample (it did not finish enough of the survey to qualify as a complete case), and six were part of the augmented sample, even though they were not technically released (four of which completed the survey).[17] Even though the remaining three practices were not part of the Frame 1 sample frame, we added them to the frame because they represented the type of practices we wanted to include in the focal population. In sum, seven of the 10 pre-test practices completed the survey. The final set of completes included all seven, with sampling weights set to 1 to indicate certainty selections. We considered the pre-test practices that did not complete the survey as eligible nonrespondents.

Exclude TINs in the Frame 1 sample that did not have a match with the OneKey file (Step 17)

Throughout Steps 8 through 17, we have presented the number of (1) TINs in the sample frame, (2) strata, (3) augmented sample TINs, and (4) release sample TINs according to the final number of TINs in the frame, strata, augmented sample TINs, and (before implementing the supplement) release sample TINs in terms of the final count. However, because we implemented these steps chronologically, sample selection occurred before we implemented a new eligibility rule, namely that at least one NPI in the TIN had to have a match with the OneKey data set. We implemented this new rule because we surmised that (1) there was an excellent chance that TINs not matching with OneKey were ineligible; or (2) TINs not matching with OneKey, even if eligible, would be extremely difficult to contact.

Therefore, the number of TINs in the Frame 1 sample frame, and in the augmented and release samples at the time of sample selection (excluding pre-test practices), were 113,160, 30,133, and 10,709, respectively, instead of 106,618, 28,529, and 10,221.[18] Moreover, at the time of sample selection, the number of strata was 248, not 236.

Divide the release sample in Frame 1 into pilot samples and the remaining sample (Step 18)

We divided the release sample into two pilots and a main release. The pilots included TINs for which we believed the contact information was superior to all other TINs; this gave us more time to do investigative

[17] All the pre-test practices that were in the frame were also part of the augmented sample. This is slightly surprising, as there were more unsampled practices in the frame than sampled practices. One of the six pretest cases (one of the four completed survey cases) in the augmented sample was actually part of the supplemental sample (Step 21), so there were two total pretest cases in the final release sample.

[18] We have previously presented counts that included the pre-test practices: 30,143 and 28,539 in the augmented sample and 10,719 and 10,231 in the release sample. These totals double-counted seven of the 10 pre-test practices in the augmented sample and one of the 10 pretest practices in the release sample.

research to improve the contact information for the remaining TINs while fielding the survey to the pilot practices. We expected the first pilot, released in July 2023, to have the best contact information, and included 1,402 TINs. The second pilot, released in August 2023, also had good contact information, though there was some uncertainty regarding cases with multiple emails. The August pilot had 882 TINs. The main release, released in October 2023, consisted of 7,943 TINs.[19]

Reselect sample cases for the release sample in Frame 1 to replace ineligible cases (Step 19)

Because 488 TINs from the Frame 1 sample were ineligible due to the eligibility criterion requiring that TINs map to OneKey, we needed to select replacement sample cases. However, we found over half the 488 TINs in one of the 12 strata with no eligible TINs. Because the budget limited us to a total sample release of 10,855 TINs, we decided to release replacement TINs only for the strata with at least one eligible TIN. As a result, we only needed to replace 226 TINs from the remaining 236 strata. Appendix D provides the allocation of replacement TINs to the remaining explicit strata. The addition of 226 TINs to the sample resulted in a new total sample size of 10,453 in Frame 1, 3 non-frame pre-test cases, and 336 in Frame 2.[20]

Select a supplemental sample from Frame 1 (Step 20)

In March 2024, we decided to add more large- and medium-sized TINs from Frame 1 to the sample in strata that included cardiology, neurology, anesthesiology, and emergency medicine in an effort to get additional survey responses from practices in these specialty areas. This included 659 large practices and 119 medium-sized practices, for a total of 778 additional practices.[21] Adding these practices to the sample resulted in a final release sample of 11,225 TINs in Frame 1 and 336 IMS_ORG_IDs in Frame 2, or 11,561 total practices. The purpose was to shore up the number of completes in these specialty areas, focusing on practices with email addresses.

The tables in Appendix D provide the counts for the allocated number of sample cases in the final release sample and the anticipated number of completed surveys allocated to each explicit stratum.

Incorporate pre-test and volunteer practices for the Frame 1 sample (Step 21)

In Step 17, we indicated that seven of the 10 pre-test practices were in the sample frame and the augmented sample, two of which were part of the final release sample. The count of 11,225 TINs in the augmented sample already included these seven cases, but we had to add the three cases that were not in the sample frame both to the frame count and to the sample cases.

Another 102 practices contacted the AMA or Mathematica to volunteer to complete the survey even though they might not have been part of the original sample. Because many of these volunteer practices were single-specialty practices for specialties with a dearth of responses, we made every effort to include

[19] The eligibility criterion applied after sampling did not affect the number of TINs in the July and August pilots. However, in the December main release, the initial number of TINs released for data collection presented here was a reduction from the actual number of TINs originally sampled in the main release: 8,394.

[20] These counts exclude 10 pre-test practices. As such, we previously presented counts that include these practices as 10,457 in the released sample.

[21] One practice in the supplemental sample was a pre-test practice.

them in the final sample of complete practices. In short, we handled these practices in a similar manner to the pre-test practices. For those that were completes, ineligibles, or eligible nonrespondents, we determined whether the practice was part of the released sample or was in the frame but not sampled. We also evaluated whether the practice was part of a related data collection effort, the CPI Survey,[22] either in the frame but not sampled or part of the CPI sample. Of the 102 volunteer practices, 52 did not respond nor could we determine their eligibility. We treated these practices the same way as we treated the single unsampled nonrespondent in the pre-test: we did not consider them when conducting survey weighting or analysis. For the remaining 50 practices, two were duplicates with practices that were already in the release sample, 34 were in the frame but not sampled, one was a new volunteer whose practice opened since 2020 and thus was not in the frame, and three were CPI practices (2 of which were sampled in the CPI sample), and 10 were found in neither the PPI nor the CPI frame. All 48 practices (excluding the duplicates with already-sampled cases) were treated as certainty cases and given a sampling weight of 1. Adding the three pre-test cases to the frame, as well as the 10 volunteers that did not match, the new volunteer, and the 3 CPI cases increased the Frame 1 total from 106,618 to 106,635. The sample total increased to 11,242.

Add relevant CPI frame cases to the focal population (Step 22)

We also augmented the previously described focal population to include some practices from the CPI frame. In particular, we added 11 clinical laboratories that indicated they had pathologists in their practice. We limited our attention to respondent laboratories because we did not know how many nonresponding laboratories included pathologists. None of these 11 laboratories matched with practices in either frame.

Survey Administration

Overview

We launched the PPI Survey in Summer 2023 and fielded it through early Fall 2024. The survey sample was released in 4 waves, the first two waves were released in Summer 2023 to plot test procedures, the third wave was released in Fall 2023, and the fourth wave was released in Spring 2024. The Physician hours survey was launched in late Fall 2023 and fielded through Summer 2024. It was released in several waves, and fielded by Mathematica, AMA and Medscape. We describe below the procedures for preparing and fielding the two surveys along with challenges encountered and solutions implemented.

Preparing to field the PPI Survey

Obtaining institutional review board approval

Before fielding the PPI Survey, we prepared and submitted our materials to the Health Media Lab Institutional Review Board (IRB). We received IRB exemption from ongoing review in accordance with the requirements of the U.S. Code of Federal Regulations for the Protection of Human Subjects,

[22] The CPI survey is based on the PPI Survey and intended to obtain information about expenses from QHPs, including chiropractors, psychologists, social workers, registered nurses, nurse practitioners, lab professionals, audiologists, speech pathologists, physical therapists, occupational therapists, optometrists, oral surgeons, podiatrists, nutritionists, and some radiologists and pathologists.

45CFR46.104(2). As we implemented changes during the field period (described later in this report) we submitted amendments to the Health Media Lab IRB and received additional approvals.

Obtaining medical specialty society endorsements

Before fielding, the AMA solicited endorsements from state medical organizations and medical organizations that represented all the physician specialties. In total, 172 state and specialty medical organizations endorsed this study. We compiled these endorsements in a document we could share with practices in our communication materials. See Appendix E for the full endorsement letter.

Updating practice contact information

Mathematica obtained practice contact information from the OneKey data set. However, this data set did not have complete contact information for every practice in our sample. Exhibit 5 provides information on the number and percent of practices in the release sample that had available contact information in the OneKey dataset. For practices missing a contact name, title, and/or email address, Mathematica and AMA staff searched the web for updated information. To do this we searched for the practice by name and address, and reviewed practice websites to gather updated contact information when possible. We searched specifically for the person at the practice who would likely be best able to complete the PPI Survey. For small physician-owned practices, this was often the physician owner or the practice manager. For larger practices, this was often the chief financial officer or vice president of finance.

In addition to this web locating, we conducted phone locating before fielding for practices that we were unable or unlikely to find information online for. For example, we found that single-physician practices were less likely to have a practice website, so web locating was challenging for those practices. For phone locating, we dialed the number we had for the practice from the OneKey data set with the goal of gathering the name, title, and email address of the person at the practice best able to respond to the PPI Survey.

Exhibit 5. Number and percent of practices in the release sample with available contact information from OneKey

	Frame 1	Frame 2	Total
Practices with contact name and email	3,930 (38%)	146 (43%)	**4,076 (38%)**
Practices with contact name but no email	4,429 (42%)	190 (57%)	**4,619 (43%)**
Practices with neither contact name nor email	2,088 (20%)	0 (0%)	**2,088 (19%)**
Total practices	**10,447**	**336**	**10,783**

We continued with web and phone locating throughout fielding, searching for updated contact information for practices if what we had proved to be incorrect (for example, if our emails were undeliverable or our letters were returned). We conducted web locating for 6,276 sampled practices and

phone locating for 4,277.[23] In total, 8,787 sampled practices (about 75 percent of all sampled practices) received either web or phone locating, or both.

Fielding the PPI Survey

We fielded the PPI Survey as an online survey in four releases. Release 1 began as a pilot in June 2023 and included the release of 1,402 practices to test the field procedures and responses by practice ownership type and specialty during the first few months. Release 2 included 1,208 additional practices released in August 2023 as part of this pilot.[24] At the end of the pilot, we reviewed the responses and discussed the need for any modifications to the survey procedures with the AMA. Modifications included the following:

- Because we saw practices clicking the link to the survey but not proceeding past the survey introduction, we shortened the text in our communication materials and the introduction so respondents could start on the actual survey sooner.
- We reviewed our procedures for releasing email communications and scheduled future releases to be at various days and times when respondents would most likely our emails.
- We increased the frequency with which we contacted practices by email and mail.

After this pilot, Release 3 included the release of 8,163 practices in October 2023. Release 4 included the release of 778 practices in March 2024. We continued fielding all four releases through September 30, 2024.

The intended respondent for the PPI Survey was the person at each practice most knowledgeable about practice finances and expenses. For small physician-owned practices, this was often the physician owner or the practice manager. For larger practices, this was often the chief financial officer or vice president of finance.

Mathematica sent practices invitation and reminder emails and letters that detailed the purpose of the survey and the reasons why participation was important, as well as other information about the survey and who at the practice should complete it. Letters and emails included the web address and a unique username and password to access the survey. Emails also included a case-specific link so respondents could click and open their individual survey directly. An outside vendor printed and mailed the invitation and reminder letters. Appendix E includes the invitation letter and email.

Communication materials also detailed the incentives offered to practices that completed the survey. We offered all participating practices a report that showed how their practice data compared with other similar practices, sent to practices when data analysis was complete. We offered small (fewer than 10 physicians) and physician-owned practices an incentive of $100 for completing the survey. Larger practices that were not physician-owned did not receive this offer because prior interviews with large practices indicated that any monetary incentive we might provide would not be meaningful to them.

[23] A total of 1,766 practices received both web *and* phone locating (for example, if web locating proved unsuccessful, we attempted phone locating).

[24] We organized the pilot into two releases based on whether we had contact information for the practice. Those that we did have contact information for we fielded in Release 1. The others we held back while we searched for contact information, then released when we had it.

However, to increase the response rate during fielding, we offered a one-year subscription to the AMA Resource-Based Relative Value Scale (RBRVS) DataManager to all practices that did not receive the monetary incentive of $100.

Our initial fielding plan involved sending practices one invitation letter and email, followed by eight reminder emails and five reminder letters spread over eight months. Practices received a communication about the survey by email or mail at least every other week, followed by up to two reminder calls during the last three months of the field period.

Challenges and solutions

Response to the PPI Survey was lower than expected. Because of this, Mathematica and the AMA made several adaptive design changes in an attempt to increase response both in general and from focal groups of practices (such as large and medium-sized practices and those with specialties with very few submitted surveys). The types of changes we made included (1) adding volunteer practices and additional releases to the sample; (2) increasing outreach to sampled practices from Mathematica, specialty societies, and the AMA; and (3) adding monetary and nonmonetary incentives for practices that completed the survey.

Added volunteer practices to the sample. Occasionally during the field period, practices that were not a part of our sample reached out to our help desk or to the AMA asking to participate in the survey. These practices typically learned about the survey from communications released by the AMA or their specialty society. We allowed these practices to participate, loading their information into our systems and sharing a survey link with them. In total we added 102 practices that requested to participate but were not a part of our initial sample to the survey. Of these, 37 (36 percent) completed the PPI Survey.

Released additional sample. We initially planned to field the PPI Survey with three sample releases. However, because of the low response, particularly from medium-sized and large practices, we added a fourth and final release of 778 medium-sized and large practices starting in March 2024 and ending in September 2024. Of these practices in the additional sample, 23 (3 percent) completed the PPI Survey.

Increased reminder outreach. The initial fielding plan involved sending reminders to practices by either mail or email every other week. However, to increase response, we revised our contact procedures to release reminder emails weekly to practices for which we had email addresses. Then, rather than continue to send reminder letters to all practices, we prioritized sending letters to practices if we did not have an email address for them, if we recently updated their contact information, or if we recently began fielding to them (as was the case with the Release 4 practices, which we began fielding to in spring 2024).

In addition to the reminder emails and letters, Mathematica conducted three rounds of reminder calls from April to June 2024, calling 2,355 practices up to three times. We prioritized phone outreach to practices in Release 4, because their field period started later than the others and they had less time with the survey; we also prioritized practices that started the survey but did not yet complete and submit it.

Specialty society outreach. Starting in May and ending in July 2024, Mathematica reached out to 33 specialty societies the AMA had indicated would be willing to contact their members to encourage them to complete the PPI Survey. After receiving IRB approval, Mathematica securely shared a list of physicians

at sampled practices within that society's specialty and asked the specialty society to reach out to these physicians and urge them to speak with their practice management colleagues to complete the survey.

AMA outreach to select practices. In June 2024, the AMA fielded the PPI Survey by email to 2,245 practices in the sample with one to five physicians that had not yet started the survey. The AMA used email addresses in its membership directory (which Mathematica did not have access to) to do this. The AMA sent the survey invitation email to a physician at each of these practices for which the AMA had an email address, assuming that at a small practice, a physician might be able to either complete the survey themselves or identify the individual at the practice who was best able to complete the survey. The AMA sent one email to these physicians, and the email contained the link to the PPI Survey hosted by Mathematica. However, none of these practices completed the PPI Survey.

Outreach to select large practices by Mathematica senior staff. To increase survey participation from larger practices (which had a lower response and were more likely to find the survey time consuming or difficult to complete given their size and the number of specialties they have), Mathematica senior staff conducted email and phone outreach to practices with more than 300 physicians who had at least clicked the survey link as of mid-April 2024. Mathematica senior staff served as a direct point of contact for these practices, offering to answer questions or provide any needed assistance in completing the survey. In total, eight of these practices (11 percent) eventually completed the PPI Survey.

Tested offering monetary incentive amounts to select practices. To increase response from medium-sized and large practices, we tested offering additional targeted incentives to a small group of practices, $500 to medium-sized practices and $1,000 to large practices. We selected these practices (10 in total) in consultation with the AMA. One of these practices ended up completing the PPI Survey, though it declined the increased incentive due to organizational policies that prevented it from accepting any survey incentive amount. Because the incentive was not effective at encouraging response, we did not expand the offer to additional practices.

Offered additional nonmonetary incentive for practices that were not small and physician-owned. To increase response from medium-sized and large practices, the AMA offered these practices a one-year subscription to the AMA RBRVS DataManager, an online resource that provides insight into how CPT and HCPCS Level II codes are created, valued, and used in the RBRVS payment system. Mathematica incorporated text about this additional incentive in communication materials to these practices.

Summary of help desk comments

Mathematica ran a help desk to field questions from practices throughout data collection. We provided a toll-free phone number and an email address where they could reach us. In total, we received 290 inquiries from practices. The most common type of inquiry we received was contact-related updates (that is, individuals at practices reaching out to let us know that we should address emails and letters to someone else at the practice)—this represented 32 percent of our PPI Survey help desk inquiries. Other common help desk inquiries included survey content-related questions (9 percent) and requests to resend survey log-in information (9 percent). We also received some inquiries about a specific practice's eligibility, the survey deadline, and requests for a deadline extension.

Physician Hours Survey

Fielding the Physician Hours Survey

We fielded the Physician Hours Survey as an online survey that ran from June 2023 to September 2024. We first fielded this survey to physicians at practices that completed the practice survey. In the practice survey, we asked respondents if they would be willing to send a link to the physician survey to all physicians in their practice. Practice survey respondents who agreed received a link to the physician survey with a message that explained the importance of the survey. Physicians at practices where the practice survey respondent did not agree to share the physician survey link became part of the physician sample that Mathematica surveyed. We did this in two releases (the first in November 2023, the second in February 2024), before fielding the survey to all physicians for whom Mathematica had email addresses in our third and final release of the physician survey in April 2024.

For the first and second releases, Mathematica contacted physicians by mail and email, sending one invitation letter and two reminder letters as well as one invitation email and three reminder emails over 10 weeks (for the first release) and seven weeks (for the second release). Release 1 fielding began the week of December 4, 2023, and ended the week of February 5, 2024. We paused outreach (and thus extended the fielding) around the holidays because we expected limited response then. Release 2 fielding began the week of February 19, 2024, and ended the week of April 1, 2024.

For the third release, Mathematica contacted physicians by email only, beginning the week of April 8 and ending the week of July 1, 2024, sending one invitation email and four reminder emails.

For all three releases, the survey remained open until August 30, 2024. Although our outreach for a specific release might have ended, physicians still had the ability to access and complete the survey until the end of August.

Six PPI Survey respondents at larger practices with complex ownership contacted the Mathematica survey team to ask if they could provide information on the hours and weeks that their physicians provided direct care, rather than fielding the separate hours survey to their physicians. In these situations we allowed them to report this information themselves within the PPI Survey and we did not field a separate survey to their physicians.

Challenges and solutions

Response to the Physician Hours Survey was generally low, in part because we had few practice survey respondents completing the PPI Survey (and even fewer sharing the link to the Physician Hours Survey with their physicians) and because the contact information we had for physicians from the OneKey data set was often inaccurate. Because of this, we decided to field the Physician Hours Survey to all physicians and use alternate contact information from the AMA and Medscape when possible. Specifically, we focused outreach to the physicians by email only (given the sample size, it was more cost effective to reach a large number of physicians by email rather than by physical mail). Details about our outreach to physicians follow.

Fielding to all physicians in the sample. To gather information on hours and weeks from more physicians within different specialties, we decided to field the Physician Hours Survey to all physicians regardless of whether their practice had already participated in the PPI Survey.

Mathematica fielded the Physician Hours Survey in April 2024 to all physicians for whom we had email addresses, excluding physicians who had already completed the hours survey, physicians at practices that reported hours and weeks in the PPI Survey, and physicians that Medscape planned to field to. This third and final release of the Physician Hours Survey by Mathematica consisted of 69,920 physicians. Mathematica conducted outreach to these physicians by email only.

Medscape fielding. The AMA coordinated with Medscape to field the Physician Hours Survey to physicians at sampled practices that were also a part of Medscape's panel of physician subscribers and for whom Medscape had an email address (60,885 in total). The AMA received separate IRB approval from the University of Illinois Chicago IRB to work with Medscape to field this survey. Mathematica shared the NPIs of physicians in sampled practices with the AMA and Medscape, and Medscape used the NPIs to identify the physicians who were also Medscape subscribers. In collaboration with the AMA, Medscape programmed a modified version of the survey using its own systems. This modified version of the Physician Hours Survey also collected practice-level contact information from select physicians in an effort to gather information that Mathematica could then use to field the PPI Survey. Specifically, the survey requested practice-level contact information from physicians in the Medscape hours survey if it was missing or determined to be incorrect by Mathematica.

AMA fielding. The AMA fielded the Physician Hours Survey to physicians for whom they had email addresses that Mathematica and Medscape did not (115,559 in total). These physicians were identified by NPI and the AMA programmed a modified version of the survey using Qualtrics. Similar to the Medscape version of the hours survey, the AMA version of the hours survey also collected practice-level contact information from all physicians willing to provide it, in an effort to gather information that Mathematica could then use to field the PPI Survey.

Summary of help desk comments

Mathematica ran a help desk to field questions from physicians throughout data collection. We provided a toll-free phone number and an email address where they could reach us. In total, we received 74 inquiries from physicians. The most common reason physicians reached out to our help desk was to indicate they were unwilling or unable to complete the survey (28 percent). Other inquiries included notification that an invite or reminder email had reached someone other than the intended recipient (that is, because the email address from OneKey was incorrect), and notifications that a physician had retired.

Response Rates and Weighting

Overview

In this section Mathematica describes two response rate calculations and the weighting methodology. When calculating both the response rate and the weights, there was some question about how to define *eligibility*. Clearly practices that were closed, merged, or duplicates were ineligible, and we considered practices located via web searches or phone calls and who refused to respond or otherwise did not

respond to be nonrespondents. However, there were practices for which we did not perform locating and that never clicked on the survey link. In these cases, we could not be sure if the practice was still open or if a physical practice location even existed.

In this study, we calculated the weights by assuming that, unless we determined practices had closed, merged, or were duplicates and therefore clearly ineligible, we considered them nonrespondents. However, we calculated the response rates in two ways. The first method matches the assumption used when calculating weights: we categorized practices that never clicked on the link as nonrespondents. The other method assumes practices that never clicked on the link were outside our focal population and, therefore, we should consider them ineligible. For both response rate calculations, the calculation of the eligibility rate did not consider practices that never clicked on the link as ineligible; it relied only on practices that were clearly eligible or ineligible due to a closure, merger, or duplication.[25]

Response rates

Response rates for the survey were based only on practices selected as part of the released probability sample. They did not include responses from the pre-test practices, volunteer practices, or clinical laboratories that reported pathology expenses in the Clinician Provider Interview (CPI) survey that were added to the analytic sample. Probability sample cases included 10,221 practices in the initial release sample from Frame 1, 1,004 practices in two supplemental samples[26] from Frame 1, and 336 practices from Frame 2, for a total of 11,561 probability sampled practices.[27] The response rate calculations did not include an additional eight pre-test practices and 102 volunteer practices. Moreover, from the CPI Survey, 11 of the 333 sampled clinical labs responded to the survey with expenses about pathologists, which we also excluded from the response rate calculations.

Exhibit 6 provides the various statuses for the 11,561 probability sample cases with the associated response rates, broken out by Frame 1, Frame 2, and overall. We also present both the weighted (using the sampling weight) and unweighted response rates, as well as eligibility rates (the estimated proportion of practices that are eligible for the survey). Because the sample did not have clustering and we selected

[25] In both response rate calculations, the denominator is the sum of completes, eligible nonrespondents, and the estimated number of eligible nonrespondents among the nonrespondents with unknown eligibility. The latter term is the product of the eligibility rate and the number of nonrespondents with unknown eligibility. In both calculations, the estimated eligibility rate is the same. However, in the first calculation the number of nonrespondents with unknown eligibility includes those who clicked on the link and those who did not, and in the second calculation the number of nonrespondents with unknown eligibility includes only those who clicked on the link.

[26] The two supplemental samples refer to (1) the sample of 226 Taxpayer Identification Numbers (TINs) meant to replace the TINs in the initial release sample that we removed because they did not match with OneKey and (2) the 778 TINs in the supplemental sample from March 2024.

[27] As noted in the sample methodology section, Frame 1 includes all TINs that had National Provider Identifiers (NPIs) that submitted Medicare claims in 2020 and were therefore in the Medicare Data on Provider Practice and Specialty (MD-PPS) 2020 data, and Frame 2 includes IMS_ORG_IDs with pediatrician NPIs that did not submit any Medicare claims in 2020.

practices with equal probability, any differences between the weighted and unweighted response rates were due to different sampling rates from stratum to stratum.[28]

Exhibit 6. Dispositions and response rates, by frame and overall

Dispositions	Frame 1	Frame 2	Overall
Total probability sample	10,221	336	11,561
Complete	323	4	327[a]
Ineligible	114	7	121
Eligible nonrespondent	261	9	270
Nonrespondent with eligibility unknown, clicked on link	4,558	242	4,800
Nonrespondents with eligibility unknown, did not click on link	5,653	53	5,706
Eligibility rate, unweighted	83.7%	65.0%	83.1%
Eligibility rate, weighted	87.9%	60.5%	83.7%
Response rate, unweighted	6.9%	2.2%	6.7%
Response rate, weighted	**7.8%**	**2.5%**	**6.8%**

[a] We included a total of 380 practice respondents in the survey analysis. In addition to the 327 practice respondents that were sampled and completed the survey, the analysis also included 53 practices that were not part of the release sample: 36 practices that volunteered to complete the survey (and did so), 6 practices that completed the PPI survey as part of the survey pretest, and 11 practices that were part of the CPI sample.

Weighting

We present the six steps used for calculating the weights:

1. Calculate practice-level sampling weights
2. Incorporate responding practices that were not part of sample as certainty selections
3. Adjust the practice-level sampling weights for two types of nonresponse (eligibility determination and cooperation among those with known eligibility)
4. Take the product of the practice-level nonresponse-adjusted weight among responding practices and the number of physicians per specialty within practice
5. Calibrate the weights to ensure weighted marginal totals match frame totals for key specialties and number of physicians per practice
6. Trim the weights to reduce the influence of outlier weights

[28] The response rate that we calculated in Exhibit 6 was equivalent to the American Association of Public Opinion Research (AAPOR) standard response rate calculation: RR AAPOR = number of completed interviews/(number of cases in the sample - estimated number of ineligible cases). We calculated this response rate by assuming that the released practices that did not click on the link for the survey were ineligible because we learned that some of these practices were either closed or ineligible for other reasons. (We surmised that, for the other practices that we had no information for, a large proportion of these practices could also have been ineligible.) However, for purposes of weighting, we assumed the practices that did not click on the link for the survey had unknown eligibility because we could not confirm that they all were ineligible. Using this assumption in the calculation of the response rates, the weighted response rates were as follows: 3.5 percent for frame 1, 2.1 percent for frame 2, and 3.4 percent overall. This version includes those who did not click on the link in the denominator, whereas the version in Exhibit 6 does not. The unweighted response rates using this assumption differed very little from the weighted version.

Calculate practice-level sampling weights (Step 1)

As described in the sampling methodology section, the sampling unit for the Physician Practice Information (PPI) Survey was the practice (the Taxpayer Identification Number [TIN] for Frame 1 and IMS_ORG_ID for Frame 2), so the initial probability weights were the inverse of the probability of selection and release for each practice.

Augmented practice sampling weight. As noted in the section on sampling methodology, we first selected an augmented sample of practices—a larger sample than we would likely need to release. Because this is an equal probability sample without clustering, the augmented sampling weight is simply the total number of practices within each stratum in the population divided by the number of practices selected in the augmented sample within that stratum. After selecting the initial augmented sample but before collecting data, the rules of eligibility changed in Frame 1 to be more restrictive, so we deemed ineligible 1,604 selected practices in the augmented sample (and 6,542 practices in the frame) .[29] As a result, we adjusted the augmented sampling weight in Frame 1 to account for the smaller number of eligible practices within each stratum. In fact, we had to remove 12 strata completely, as there were no eligible practices in those strata based on the revised rules. (Appendix C provides listings of the strata and their population counts.) We selected 29,539 practices in the augmented sample (28,539 from Frame 1 and 1,000 from Frame 2), of which 11,561 were in the release probability sample. We also identified an additional 40 as part of the frame and released them as pre-test or volunteer cases, leaving 17,978 that we never released for data collection. After applying the new eligibility criterion, the total number of practices in Frame 1 was 106,618 and 14,771 in Frame 2, for a total of 121,389 practices, in which the definition of *practice* depended on the frame. The sum of the augmented sampling weights necessarily was equal to this total number.

Initial release practice sampling weight. The initial release sampling weight was the weight applied to augmented sample practices that we initially released for data collection. It had two components: the augmented sampling weight and an adjustment to the augmented sampling weight to account for cases released from the augmented sample. The number of TINs (for Frame 1) and IMS_ORG_IDs (for Frame 2) in the initial release sample depended on an expected completion rate of 30 percent. We considered TINs and IMS_ORG_IDs that were not part of the initial release sample part of the reserve sample. For Frame 1, we separated the initial release sample into two pilots in July and August 2023, and a main release in December 2023; we released all initial release sample cases for Frame 2 in August 2023. (We discuss adding more release cases in supplemental samples selected from the reserve sample later in this section.) The separation of the Frame 1 initial release sample into these three segments was not random, as the pilots consisted primarily of TINs with the best contact information, but we eventually released all TINs in the initial release sample for data collection.

[29] In the Sample frame development and sample selection overview portion of the sampling methodology section, we lay out 22 steps for creating the sample frame and selecting the sample. Step 8 includes an eligibility criterion added after we completed the initial sample frame creation and sample selection: we required the TINs in Frame 1 to have a match with the OneKey data set because we surmised that either (1) the practices had closed between 2020 and 2023 and would be ineligible anyway or (2) their lack of contact information from OneKey would have made the open practices extremely difficult to contact. This reduced the number of eligible TINs in Frame 1 and it reduced the size of the augmented sample selected from that frame.

In the Frame 1 sampling strata when the sampling fraction was less than one (that is, not all TINs were selected), we created two substrata per stratum. The two substrata separated TINs with relatively good contact information from those that did not have good contact information. We updated the sample allocation so that, in strata in which we did not select all practices, the sampling rate in the good-contact substrata was much higher than that of the TINs with poor contact information.[30] Because the sampling weight was the inverse of the probability of selection and there were different sampling rates for these two types of substrata, the sampling weights for TINs with poor contact information were necessarily greater than those with good contact information. With the changes in eligibility rules,[31] the adjustments to the augmented sampling weight also affected the initial release sampling weight. In fact, we deemed 488 TINs in the initial release sample ineligible and did not release them for data collection.

Final release practice sampling weight. Completion rates in the PPI Survey were lower than expected, particularly so for some specialties. To account for the 488 ineligible TINs in the initial release sample, we selected and released 226 replacement TINs from strata with underperforming specialties. In March 2024, we also released a supplemental sample to large and medium TINs from Frame 1 in select strata. We focused on strata that included cardiology, neurology, anesthesiology, and emergency medicine to get additional survey responses from practices in these specialty areas. We selected both supplemental samples from the reserve sample within single-specialty strata corresponding to these four specialties and without regard to the contact information substrata.[32] Therefore, the final release sampling weights had to account for all chances of selection from the augmented sample so the sum of the final release weights added up to 121,389.

Incorporate responding practices that were not part of the release probability sample (Step 2)

As stated in the sample methodology section, several responding practices included in the final analytic sample were not part of the release sample. The very low completion rate in the release sample persuaded us to incorporate as many of these non-sampled responding practices as feasible. Additional practices in the final analytic sample included seven pre-test practices, 35 volunteer practices, and 11 clinical laboratories from the CPI sample practices, all of which completed the survey, giving us a total of 380 completed surveys. Details for each of these extra completes follow.

Pre-test practices. Ten pre-test practices agreed to complete the PPI Survey and provide feedback to the American Medical Association (AMA) and Mathematica. As noted in the sample methodology section, these 10 pre-test practices were not selected from either frame; they volunteered after solicitation by the AMA and Mathematica. Nevertheless, we matched seven of the pre-test practices to TINs in Frame 1 and none to IMS_ORG_IDs in Frame 2; interestingly, we selected all seven of the matched TINs as part of the augmented sample: one was in the initial release sample and six were in the reserve sample, one of which was selected for the supplemental sample in March 2024. The single TIN in the initial release sample did

[30] In fact, we released all cases from the augmented sample in the good contact substratum for data collection, so the release adjustment for those substrata was equal to one.

[31] The sample frame development and sample selection overview portion of the sampling methodology section describes the changes in the eligibility criteria, as noted in a previous footnote.

[32] We selected all available TINs that were in the good contact information substrata in the initial release sample, so none were available for these supplemental samples.

not complete enough of the survey to qualify as a completed survey and was treated as an eligible nonrespondent. Of the six remaining TINs in the reserve sample, both the single TIN in the supplemental sample and three others were respondents, and two were eligible nonrespondents. The two TINs in the final release sample were given their selected sampling weight; the remaining five TINs were treated as certainty cases with a sample weight of one. The weights of the eligible nonrespondents were appropriately reallocated to other practices in the nonresponse adjustment phase.

We added the three TINs that were not part of the Frame 1 sampling frame to both the frame and the final analytic sample because they represented the type of practices we wanted to include in the focal population. All three completed the survey. This increased the Frame 1 total from 106,618 to 106,621. These practices also had a sampling weight of one.

Volunteer practices. Another 102 practices contacted the AMA or Mathematica to volunteer to complete the survey even though they might not have been part of the original sample. As noted in the sample methodology section, we handled these practices in a similar manner to the pre-test practices, trying to match them to the existing frames, adding respondents to the sample (and frame) that were not matched to the frame, and dropping nonresponding volunteers. Exhibit 7 summarizes the status of each of the volunteers.

Exhibit 7. Status of volunteer practices

Volunteer type	Number of cases	Resolution	Instrument statuses	Sampling weight of volunteer type
Total volunteers	102	.	.	
Matched with TIN in release sample	2	Treated as duplicates; only 1 case kept in sample for each pair	1 complete, 1 eligible nonrespondent	Sample weight of case in release sample
Matched with TIN in sample Frame 1	34	Added to release sample	25 completes, 7 eligible nonrespondents, 2 ineligibles	1
New TIN since 2020	1	Added to release sample	1 complete	1
Not a new TIN, did not match with either sample frame	10	Added to release sample	8 completes, 1 eligible nonrespondent, 1 ineligible	1
Matched with TIN in CPI release sample	2	Added to release sample	1 complete, 1 eligible nonrespondent	1
Matched with TIN in CPI frame	1	Dropped	1 nonrespondent, eligibility unknown	0
Eligibility unknown, did not attempt to match to PPI nor CPI frame	52	Dropped	52 nonrespondents, eligibility unknown	0

CPI sample practices. All 11 clinical laboratories that provided useable information in the CPI Survey provided expense data for pathologists. We confirmed that these 11 practices did not match with any practices in either PPI frame and we added them as certainty cases.

Adjust the practice-level sampling weights for two types of nonresponse (eligibility determination and cooperation among those eligible) (Step 3)

We used weighting classes to adjust the practice weights for nonresponse. We accomplished this by adjusting for two types of nonresponse: ability or inability to determine if a practice is an eligible nonrespondent among practices known to be eligible. We calculated each adjustment by forming classes of practices with similar characteristics and then using the inverse of the class response rate as the adjustment factor in that class. The adjusted weight is the product of the base weight and the adjustment factors. We treated ineligible practices as eligibility known in the first stage of adjustment, then dropped them before computing adjustments in the second stage. We constructed weighting classes with sufficient counts in each class to make the adjustment more stable (that is, to minimize their variability). When forming the weighting classes, it was important to control for specialties and the facility billing status as much as possible. In Exhibit 1 of the sampling methodology section, we provided a crosswalk of the 65 specialties as defined by Medicare collapsed to 47 specialties considered important for analysis. The AMA in turn collapsed these to 39 specialties for sampling. To address the lower response rate for some specialties and subspecialties, the specialty category was collapsed to 19 less granular categories that were otherwise described by the Medicare specialty. The rolling up to a less granular category was accomplished by grouping "like" specialties together, as defined by historically similar practice expense by hour and typical site-of-service. For example, primary care represents family medicine, internal medicine, pediatric medicine and geriatric medicine.

Exhibit 8 shows the crosswalk for the 39 specialties collapsed to 19.

Exhibit 8. Crosswalk of 19 specialties used for weighting to 47 specialties for analysis and 39 specialties for sampling

Grouped 47 practice sampling specialties	Grouped 39 stratum specialties	Grouped 19 specialties for weighting
Cardiac Electrophysiology	Cardiology	Cardiology
Cardiology	Cardiology	Cardiology
Interventional Cardiology	Interventional Cardiology	Cardiology
Dermatology	Dermatology	Dermatology
Gastroenterology	Gastroenterology	Gastroenterology
Hematology/Oncology	Hematology/Oncology	Hematology/Oncology
Anesthesiology	Anesthesiology	Hospital Based Medicine
Critical Care Medicine	Hospital Based	Hospital Based Medicine
Emergency Medicine	Emergency Medicine	Hospital Based Medicine
Hospice & Palliative Medicine	Hospital Based	Hospital Based Medicine
Hospitalist	Hospital Based	Hospital Based Medicine
Infectious Disease	Hospital Based	Hospital Based Medicine
Cardiothoracic Surgery	Cardiothoracic Surgery	Hospital Based Surgery
Colorectal Surgery	General Surgery	Hospital Based Surgery
General Surgery	General Surgery	Hospital Based Surgery
Neurosurgery	Neurosurgery	Hospital Based Surgery
Obstetrics/Gynecology	Obstetrics/Gynecology	Obstetrics/Gynecology

Grouped 47 practice sampling specialties	Grouped 39 stratum specialties	Grouped 19 specialties for weighting
Endocrinology	Endocrinology	Office Based Medicine
Rheumatology	Rheumatology	Office Based Medicine
Nephrology	Nephrology	Office Based Medicine
Neurology	Neurology	Office Based Medicine
Pain Medicine	Pain Medicine	Office Based Medicine
Physical Medicine & Rehabilitation	Physical Medicine & Rehabilitation	Office Based Medicine
Allergy & Immunology	Allergy & Immunology	Office Based Proceduralist
Interventional Pain Medicine	Interventional Pain Medicine	Office Based Proceduralist
Plastic Surgery	Plastic Surgery	Office Based Proceduralist
Sleep Medicine	Sleep Medicine	Office Based Proceduralist
Urology	Urology	Office Based Proceduralist
Ophthalmology	Ophthalmology	Ophthalmology
Orthopedic Surgery	Orthopedic Surgery	Orthopedic Surgery
Hand Surgery	Hand Surgery	Orthopedic Surgery
Otolaryngology	Otolaryngology	Otolaryngology
Pathology	Pathology	Pathology
Family Medicine	Family Medicine	Primary Care
Sports Medicine	Family Medicine	Primary Care
General Practice	General Practice	Primary Care
Geriatric Medicine	Geriatric Medicine	Primary Care
Internal Medicine	Internal Medicine	Primary Care
Osteopathic Manipulative Medicine	Internal Medicine	Primary Care
Pediatrics	Pediatrics	Primary Care
Interventional Radiology	Interventional Radiology	Radiology
Psychiatry	Psychiatry	Psychiatry
Pulmonary Disease	Pulmonary Disease	Pulmonary Disease
Radiation Oncology	Radiation Oncology	Radiation Oncology
Nuclear Medicine	Radiology	Radiology
Radiology	Radiology	Radiology
Vascular Surgery	Vascular Surgery	Vascular Surgery

Eligibility determination step. Exhibit 9 shows the 41 weighting classes used in the eligibility determination step, complete with the number of practices in each weighting class, the number of practices with eligibility known, and the maximum adjustment in each class. If we had not created weighting classes and did only a straight ratio adjustment across all practices, the size of the adjustment would be 16.33; we tried to constrain the size of the adjustment to twice this number. Three of the 41 classes exceeded this threshold. The maximum adjustment overall was 41.7 in the weighting class of single specialty hospital-based medicine practices with least 50 percent or unknown facility billing. We also attempted to require at least 10 responding and 20 total practices in a weighting class; however, given the priority of maintaining separate weighting classes for the 19 collapsed specialties in single specialty

practices and the very low response rates, we could not meet this goal in five of the 41 weighting classes. Three of the classes had five or fewer responding practices (hematology/oncology single specialty, pulmonary disease, and radiation oncology).

Exhibit 9. Weighting classes for eligibility determination step, with number of practices in each class and number of practices with eligibility known

Weighting class	Count of practices	Count of practices with known eligibility	Maximum weighting adjustment
Cardiology single specialty, facility billing under 25% or unknown	181	10	21.4
Cardiology single specialty, facility billing at least 25%	199	11	14.0
Dermatology single specialty	319	27	11.7
Gastroenterology single specialty	193	13	15.5
Hematology/Oncology single specialty	81	4	27.5
Hospital Based Medicine single specialty, under 50% facility billing	143	9	15.2
Hospital Based Medicine single specialty, at least 50% or unknown facility billing	645	20	41.7
Hospital Based Surgery single specialty, under 75% facility billing	166	16	10.9
Hospital Based Surgery single specialty, at least 75% facility billing	267	13	16.4
Multiple specialty, at least 75% primary care specialties	277	18	24.6
Multiple specialty, under 50% or unknown facility billing, less than 2 orthopedic surgeons	643	25	30.5
Multiple specialty, under 50% or unknown facility billing, 2 or more orthopedic surgeons	193	10	24.2
Multiple specialty, at least 50% facility billing, fewer than 2 orthopedic surgeons	582	23	33.4
Multiple specialty, at least 50% facility billing, 2 or more orthopedic surgeons	364	32	18.1
Obstetrics/gynecology single specialty, under 25% facility billing	357	29	11.9
Obstetrics/gynecology single specialty, at least 25% facility billing	124	10	10.7
Office Based Medicine single specialty, under 25% or unknown facility billing	527	27	18.3
Office Based Medicine single specialty, 25% to under 50% facility billing	148	13	12.1
Office Based Medicine single specialty, at least 50% facility billing	306	11	35.8
Office Based Proceduralist, under 25% or unknown facility billing	340	26	15.8
Office Based Proceduralist, at least 25% facility billing	308	16	15.3
Ophthalmologist, under 25% or unknown facility billing	288	24	12.9
Ophthalmologist, at least 25% facility billing	225	18	12.4
Orthopedic Surgery, under 50% facility billing	248	17	22.7
Orthopedic Surgery, at least 50% facility billing	202	16	13.9
Otolaryngology single specialty	182	16	11.5
Pathology single specialty	72	27	11.5
Pediatrics Frame 2 single specialty	336	20	16.9

Weighting class	Count of practices	Count of practices with known eligibility	Maximum weighting adjustment
Primary Care, under 25% or unknown facility billing, physician owned	1,035	72	15.2
Primary Care, under 25% or unknown facility billing, corporate owned not vertically integrated	373	34	10.2
Primary Care, under 25% or unknown facility billing, corporate owned, vertically integrated	528	38	13.0
Primary Care, 25% to less than 50% facility billing	228	13	15.7
Primary Care, 50% to less than 75% facility billing	142	10	15.9
Primary Care, 75% to 100% facility billing	219	8	30.7
Psychiatry, under 25% or unknown facility billing	344	24	14.2
Psychiatry, at least 25% facility billing	273	12	29.8
Pulmonary Disease	107	5	20.2
Radiation Oncology	58	2	19.0
Radiology, under 25% or unknown facility billing	149	21	18.0
Radiology, at least 25% facility billing	172	38	8.7
Vascular Surgery	88	10	7.7
TOTAL/MAXIMUM	**11,632**	**788**	**41.7**

Response among the eligible step. Next, we dropped ineligible practices and calculated another adjustment to account for respondents among those known to be eligible. Exhibit 10 shows the 27 weighting classes used in this step, complete with the number of practices in each weighting class, the number of responding practices, and the maximum adjustment in each class. If we had not created weighting classes and did only a straight ratio adjustment across all practices, the size of the adjustment would be 1.82; we tried to constrain the size of the adjustment to twice this number. One of the 27 classes exceeded this threshold. The maximum adjustment overall was 3.75 in the weighting class of single-specialty hematology/oncology practices. We tried to maximize the number of practices in each weighting class but had five or fewer practices in five of the 27 classes.

Exhibit 10. Weighting classes for response among eligible practices step, with number of practices in each class and number of responding practices

Weighting class	Count of eligible practices	Count of responding practices	Maximum weighting adjustment
Cardiology single specialty	18	11	1.72
Dermatology single specialty	26	17	1.40
Gastroenterology single specialty	10	5	2.29
Hematology/Oncology single specialty	4	2	3.75
Hospital Based Medicine single specialty	21	10	2.06
Hospital Based Surgery single specialty	20	12	1.44
Multiple specialty, at least 75% primary care specialties	11	5	2.46

Weighting class	Count of eligible practices	Count of responding practices	Maximum weighting adjustment
Multiple specialty, under 50% or unknown facility billing, less than 2 orthopedic surgeons	19	13	1.72
Multiple specialty, at least 50% facility billing, less than 2 orthopedic surgeons	16	10	1.77
Multiple specialty, 2 or more orthopedic surgeons	36	18	2.39
Obstetrics/gynecology single specialty	35	19	1.68
Office Based Medicine single specialty, under 25% or unknown facility billing	25	15	1.59
Office Based Medicine single specialty, at least 25% facility billing	20	9	1.66
Office Based Proceduralist single specialty	38	19	2.25
Ophthalmologist single specialty	37	25	1.53
Orthopedic Surgery single specialty	25	15	1.32
Otolaryngology single specialty	13	7	1.63
Pathology single specialty	24	19	1.55
Pediatrics frame 2 single specialty	13	4	2.79
Primary Care, under 25% or unknown facility billing	126	74	1.74
Primary Care, at least 25% facility billing	29	10	2.42
Psychiatry single specialty	26	12	2.11
Pulmonary Disease single specialty	4	3	1.42
Radiation Oncology single specialty	2	2	1.00
Radiology, under 75% facility billing	18	9	2.23
Radiology, at least 75% or unknown facility billing	37	29	1.37
Vascular Surgery	8	6	1.34
TOTAL/MAXIMUM	**661**	**380**	**3.75**

Take the product of the practice-level nonresponse-adjusted weight among responding practices and the number of physicians per specialty within each practice (Step 4)

Although the sampling unit is the practice, the unit of analysis is physician-level expenses per hour. From the survey, we obtained expenses by physician specialty within each practice. Each line of the resulting data set corresponded to a practice–specialty combination—a "department"—in the responding practice. To create a weight for each line of data that accounted for physician-level expenses, we took the product of the number of survey-reported physicians and the nonresponse-adjusted practice weight for each specialty recorded for a responding practice. When summing this product specialty weight across all departments, the total number of physicians was estimated to be 2,056,784, nearly three times larger than the total number of physicians from the two sample frames, which was 693,502.

A review of these data indicated a large potential for measurement error in this estimate. For example, some practices entered a value for the number of physicians per specialty that was likely much higher than the actual value. One of the responding practices (corresponding to a TIN from MD-PPAS) reported that they had 2,870 physicians, including 2,041 anesthesiologists, even though the sample frame (MD-PPAS) indicated that it only had 102 physicians. The result was that the responses corresponding to this practice had huge weights, and the counts of reported physicians for practices with a similar discrepancy to the sample frame are dubious.[33] To alleviate this source of error, we flagged practices that reported at least 10 physicians in their practice and when that count was more than double the number of physicians recorded in the sample frame.[34] This occurred in 19 of the 380 responding practices and in 33 of the 831 departments. In these cases, we replaced the total count entered by the responding practice with the sample frame value and then performed a ratio adjustment of the product specialty weights for each department to add up to the frame total. When summing this adjusted product specialty weight across all departments, the total number of physicians was estimated to be 982,355, which, although still higher than the total number of physicians from the two sample frames, is much closer to the value of 693,502.

Calibrate the weights to ensure the weighted marginal totals match frame totals for key specialties, number of physicians per practice, and other key variables when possible (Step 5)

The sum of the product specialty weights described in step 4 provided an estimate of the total number of physicians but had high variance because of the low response rate. We therefore performed a ratio adjustment of both versions of the product specialty weights described in step 4 so that they summed to 693,502, the total number of physicians in the two frames.

For some specialties, the sum of the product specialty weights still deviated substantially from the frame totals, particularly when the product specialty weight did not include the extra adjustment, indicating a potential for bias if these weights are not further adjusted. In addition, the highest concentration of physicians in the responding practices were from large practices, which deviated substantially from the population. For example, 281 of the 289 pediatricians from responding practices came from practices with more than 100 physicians, whereas in the population, less than 2 percent of pediatricians come from such large practices. To alleviate these biases, we calibrated the weights using a raking procedure[35] to ensure the marginal weighted sum of the number of physicians per specialty and for categories of practice size

[33] It is possible, but highly unlikely, that there were 2,768 physicians that either did not charge Medicare or were new since 2020 in that TIN. It's also possible that the survey respondent answered for more than one TIN within their organization, and the Medicare Data on Provider Practice and Specialty result was only for one TIN. If that was the case, then the weights were too large anyway, because the survey response would represent multiple TINs, not just the selected one.

[34] The minimum of 10 was used because it is not uncommon for the sample frame to indicate, for example, one physician in the practice but for the survey report to indicate two physicians. For these small differences, it seemed better to believe the survey report, and any adjustment would have little impact on the overall estimate of the total number of physicians. The minimum of "double the number in the sample frame" was used because, again, it was also not uncommon for there to be small differences even for larger TINs for which believing the survey report was defensible.

[35] Iterative proportional fitting, or raking, is a method of adjusting weights in an iterative sequential manner so weighted marginal totals on key variables of interest match those of the population one variable at a time. Statisticians consider it a type of post-stratification.

matched population totals for select variables. We did this for both versions of the weights (product specialty and adjusted product specialty that were created in step 4).

Exhibit 11 presents the *specialty categories* we used for raking. For ease of presentation, we abbreviated some of the broad category labels. This exhibit shows the responding practice specialties (column A), the count of departments from responding practices (column B), and the number of physicians reported in the survey by specialty in column C. We show the summed totals for the product specialty weights (both unadjusted and adjusted) in columns D and E respectively. Column F shows the distribution of physicians by specialty based on the sample frame counts. The ratios in columns G and H are not the actual raking adjustments. Instead, they show how the sums of the product of the physician count and practice weight in columns D and E (respectively) differ from the actual population count in column F for each specialty. The adjustments presented here show how the weighted totals vary from the population *before* raking.

For most specialties, the specialty categories that we used for this raking step, shown in Exhibit 11, matched the 19 broad specialty categories provided in Exhibit 8. For certain broad specialties, however, we took an additional step to control for smaller groups of specialties within the broad category by percentage of Medicare billing. We did this by combining specialties with similar percentages of Medicare billing in a facility setting among the constituent physicians. We aimed to include at least 15 departments per category, but this was not possible for all categories. For example, the facility billing percentage profile for allergists and immunologists differed from other specialists in the office-based proceduralist category, so, therefore, despite the small number of departments, we did not combine them with another category.

Exhibit 11. Calibration of counts of physicians by specialty

A	B	C	D	E	F	G	H
Specialty	Count of specialty–practice combinations in sample	Count of physicians in sample[a]	Summed total of physician-level unadjusted specialty weights	Summed total of physician-level specialty weights, adjusted for frame-survey differences	Population total of physicians from frame to be matched	Ratio of population total and summed unadjusted specialty weight	Ratio of population total and summed adjusted specialty weight
Total	831	18,086	693,502	693,502	693,502		
Cardiology	30	523	10,399.34	21,767.97	27,429	2.64	1.26
Dermatology	29	207	4,934.87	10,329.71	12,501	2.53	1.21
Gastroenterology	20	320	7,031.85	14,719.13	14,622	2.08	0.99
HBM: Hospitalist, Emergency Medicine, Infectious Disease	36	4,182	256,906.3	85,049.22	97,433	0.38	1.15
HBM: Anesthesiology, Critical Care Medicine, Hospice & Palliative Medicine	23	990	46,860.1	58,851.90	47,893	1.02	0.81

A	B	C	D	E	F	G	H
Specialty	Count of specialty–practice combinations in sample	Count of physicians in sample[a]	Summed total of physician-level unadjusted specialty weights	Summed total of physician-level specialty weights, adjusted for frame-survey differences	Population total of physicians from frame to be matched	Ratio of population total and summed unadjusted specialty weight	Ratio of population total and summed adjusted specialty weight
HBS: Cardiothoracic and Neurosurgery	29	271	6,209.79	12,998.37	9,095	1.46	0.70
HBS: General and Colorectal Surgery	39	517	15,130.82	31,671.95	25,560	1.69	0.81
Hematology/ Oncology	19	482	9,785.27	20,482.60	13,945	1.43	0.68
OBM: Neurology, Nephrology, Physical Medicine & Rehabilitation	44	821	18,319.68	37,310.57	33,636	1.84	0.90
OBM: Endocrinology, Rheumatology, Pain Medicine	59	1,045	72,315.03	23,990.61	14,104	0.20	0.59
OBP: Plastic Surgery, Sleep Medicine, Urology	42	286	7,464.29	12,594.32	15,399	2.06	1.22
OBP: Interventional Pain Medicine, Allergy & Immunology	20	523	8,849.05	18,216.09	5,183	0.59	0.28
Obstetrics/ gynecology	31	408	10,972.4	22,967.52	33,481	3.05	1.46
Ophthalmology	33	295	8,549.69	17,896.28	17,992	2.10	1.01
Orthopaedic Surgery	34	486	16,372.09	26,661.15	24,410	1.49	0.92
Otolaryngology	21	236	7,964.33	8,448.41	9,611	1.21	1.14
PC: Internal Medicine	44	624	38,167.93	29,436.89	76,481	1.90	2.60
PC: Pediatrics and Geriatrics	26	391	14,145.36	29,609.18	38,462	2.72	1.30
PC: All others	79	1,333	45,121.82	75,681.86	93,180	2.07	1.23
Pathology	30	537	6,729.88	13,963.92	3,309	0.49	0.24
Psychiatry	24	501	16,371.95	20,458.67	26,447	1.62	1.29
Pulmonary Disease	21	341	9,330.42	15,261.47	10,455	1.12	0.68
Radiation Oncology	14	227	6,583.86	13,781.38	4,931	0.76	0.36

A	B	C	D	E	F	G	H
Specialty	Count of specialty–practice combinations in sample	Count of physicians in sample[a]	Summed total of physician-level unadjusted specialty weights	Summed total of physician-level specialty weights, adjusted for frame-survey differences	Population total of physicians from frame to be matched	Ratio of population total and summed unadjusted specialty weight	Ratio of population total and summed adjusted specialty weight
Radiology: Diagnostic Radiology	39	1,841	36,153.93	52,210.21	31,511	0.87	0.60
Radiology: Interventional Radiology, Nuclear Medicine	25	585	10,133.03	14,960.40	2,723	0.27	0.18
Vascular Surgery	20	114	2,698.90	4,182.23	3,709	1.37	0.89

[a] The count of physicians in the sample is based on the respondents' entries of the count of physicians per specialty. It does not include the adjustment that reduces the count of physicians because of large deviations from the frame.

HBM = hospital-based medicine; HBS = hospital-based surgery; OBM = office-based medicine; OBS = office-based surgery; PC = primary care.

Exhibit 12 shows the *practice size categories* we used for raking, with responding practices organized by practice size as denoted by number of physicians in the practice. Column A shows the practice size by the number of physicians in the practice across all specialties. This number is derived from the original values specified in the frame, not the counts provided by the respondents. Column B shows the number of departments in responding practices of the size indicated in column A. For example, 316 departments occurred in practices that had, across all specialties, one to 10 physicians. The total number of physicians reported in the survey in practices in each size category is shown in column C. For example, the 316 departments reported a total of 1,928 physicians in the survey. The summed totals for the product specialty weights (both unadjusted and adjusted) are shown in columns D and E respectively. Column F shows the distribution of physicians by practice size based on the sample frame counts. As with Exhibit 11, the ratios in columns G and H are not the actual raking adjustments but instead show how the sums of the product of the physician count and practice weight in columns D and E (respectively) differ from the actual population count in column F for each practice size category. The adjustments presented here show how the weighted totals vary from the population *before* raking.

The concentration of physicians in the responding practices come from larger practices compared with those in the frame. Note that the population totals in column F were scaled to the total number of unique physicians. The actual population counts were of practice–physician combinations, of which there were 890,819, a higher number because some physicians were affiliated with more than one practice.

Exhibit 12. Calibration of counts of physicians by practice size

A	B	C	D	E	F	G	H
Practice size by number of physicians[a]	Count of specialty–practice combinations in sample	Count of physicians in sample[b]	Summed total of physician-level unadjusted specialty weights	Summed total of physician-level specialty weights, adjusted for frame-survey differences	Population total of physicians from frame to be matched	Ratio of population total and summed unadjusted specialty weight	Ratio of population total and summed adjusted specialty weight
Total	831	18,086	693,502	693,502	693,502		
1 to 10	316	1,928	183,954.9	129,597.6	169,215.4	0.92	1.31
11 to 100	82	1,950	152,538.5	92,609.6	202,509.5	1.33	2.19
101 and over	433	14,208	357,008.7	471,294.8	321.777.1	0.90	0.68

HBM = hospital-based medicine; HBS = hospital-based surgery; OBM = office-based medicine; OBS = office-based surgery; PC = primary care.

[a] The practice sizes within the practice-size categories are based on the practice sizes from the frame, not from the sample. Therefore, there is no clear pattern showing that the responding practices are always larger. Some of the practices in the smaller categories were among the 33 practice–specialty combinations that indicated much larger practice sizes. When doing the raking, the weights being raked are based on these frame-based practice sizes among responding practices.

[b] The count of physicians in the sample is based on the respondents' entries of the count of physicians per specialty. It does not include the adjustment that reduces the count of physicians because of large deviations from the frame.

The actual raking adjustments (post-raking) ranged from 0.12 to 12.54 for the unadjusted specialty weights and from 0.11 to 7.04 for the adjusted specialty weights that were adjusted for large survey-frame deviations.

Trim the weights to reduce the influence of outlier weights (Step 6)

Once we made the raking adjustments, we assessed the distribution of the adjusted weights for unusually high values, which could make the survey estimates less precise. We used the design effect attributed to the variation in the sampling weights as a statistical measure to determine both the need for and amount of trimming. The design effect attributed to weighting is a measure of the potential loss in precision caused by the variation in the sampling weights relative to a sample of the same size with equal weights. We also wanted to minimize the extent of trimming to avoid the potential for bias in the survey estimates. Therefore, the decision to trim required us to balance increasing bias and decreasing variance.

We conducted the trimming within 26 trimming classes defined by the 26 specialty groups shown in Exhibit 11. Exhibit 13 shows the design effects attributable to unequal weights before and after trimming, within trimming classes defined by those 26 specialty groups for both the unadjusted and adjusted weights.

Exhibit 13. Design effects before and after trimming within the 26 specialty groups

Specialty	Count of specialty-practice combinations in sample	Number of cases trimmed, unadjusted weights	Design effect attributed to unequal unadjusted weights		Number of cases trimmed, adjusted weights	Design effect attributed to unequal adjusted weights	
			Before trimming	After trimming		Before trimming	After trimming
Total	831	27			28		
Cardiology	30	1	2.50	2.47	0	2.34	2.34
Dermatology	29	1	3.42	2.46	1	3.20	2.07
Gastroenterology	20	0	1.90	1.90	0	2.11	2.11
HBM: Hospitalist, Emergency Medicine, Infectious Disease	36	2	11.55	5.56	2	6.72	3.26
HBM: Anesthesiology, Critical Care Medicine, Hospice & Palliative Medicine	23	2	4.46	4.00	2	6.82	3.29
HBS: Cardiothoracic and Neurosurgery	29	0	1.98	1.98	1	2.63	2.18
HBS: General and Colorectal Surgery	39	0	2.01	2.01	0	1.97	1.97
Hematology/ Oncology	19	1	3.89	2.39	1	4.98	2.36
OBM: Neurology, Nephrology, Physical Medicine & Rehabilitation	44	0	1.98	1.98	1	2.21	2.10
OBM: Endocrinology, Rheumatology, Pain Medicine	59	1	22.28	4.85	2	4.68	2.54
OBP: Plastic Surgery, Sleep Medicine, Urology	42	1	2.38	1.96	0	1.94	1.94
OBP: Interventional Pain Medicine, Allergy & Immunology	20	0	2.71	2.71	0	2.68	2.68
Obstetrics/ gynecology	31	0	2.14	2.14	0	1.81	1.81
Ophthalmology	33	2	5.05	2.53	2	4.62	2.10
Orthopaedic Surgery	34	1	9.15	3.82	3	5.30	3.61
Otolaryngology	21	1	4.53	2.28	0	1.50	1.50

Specialty	Count of specialty-practice combinations in sample	Number of cases trimmed, unadjusted weights	Design effect attributed to unequal unadjusted weights		Number of cases trimmed, adjusted weights	Design effect attributed to unequal adjusted weights	
			Before trimming	After trimming		Before trimming	After trimming
PC: Internal Medicine	44	3	4.17	3.16	2	2.87	2.22
PC: Pediatrics and Geriatrics	26	1	12.69	3.48	1	2.05	2.04
PC: All others	79	1	2.70	2.43	1	3.69	2.67
Pathology	30	2	3.64	3.03	2	3.28	2.97
Psychiatry	24	2	3.48	2.87	1	3.46	2.37
Pulmonary Disease	21	1	4.00	2.28	1	6.10	2.39
Radiation Oncology	14	1	8.21	3.18	1	9.00	3.42
Radiology: Diagnostic Radiology	39	1	2.79	2.68	1	3.15	2.85
Radiology: Interventional Radiology, Nuclear Medicine	25	1	4.16	3.50	2	4.43	3.33
Vascular Surgery	20	1	2.00	1.82	1	2.73	1.82

[a] The count of physicians in the sample is based on the respondents' entries of the count of physicians per specialty. It does not include the adjustment that reduces the count of physicians because of large deviations from the frame.

HBM = hospital-based medicine; HBS = hospital-based surgery; OBM = office-based medicine; OBS = office-based surgery; PC = primary care.

Trimming substantially reduced the design effect for some of the trimming classes for both the unadjusted and adjusted weights. For example, for the trimming class "OBM: Endocrinology, Rheumatology, Pain Medicine" the design effect due to unequal weighting among the unadjusted weights was reduced from 22.28 to 4.85, and among the adjusted weights this design effect was reduced from 4.68 to 2.54.

Analysis

Overview

Our goal was to calculate average practice expense per hour of patient care provided at the department level. We calculated mean practice expense using the data provided by practices that completed the PPI Survey, and we calculated the mean number of patient care hours provided by different physician specialties using data provided by physicians in the Physician Hours Survey. In this section, we first describe the steps taken to analyze the data from the Physician Hours Survey and then describe steps taken to analyze the PPI Survey. The AMA conducted this analysis with input from Mathematica.

Physician Hours Survey data

We took the following steps to analyze the Physician Hours Survey data at the physician level:

Excluded missing data (Step 1). We excluded observations in which physicians reported that they spent zero hours and weeks providing patients care in 2022 or left these questions blank.

Grouped physician data by specialty (Step 2). We grouped physician data into the 19 rolled-up specialties listed in Exhibit 7 based on their reported Medicare specialty.

Created indicator variable for part time or full time (Step 3). We segmented the physician hours data by generating an indicator variable for part-time and full-time work using the information that each physician reported in question five of the survey. We defined part time as physicians who worked less than 35 hours a week and full time as physicians who worked 35 or more hours a week. This was because the density plot shows that the most common reported number of minimum hours reported by practices in the practice survey for full-time physicians is around 35 hours per week.

Computed physicians' mean hours and weeks worked (Step 4). We computed mean hours and mean weeks separately for part-time and full-time physicians for each of the 19 specialties.

PPI Survey data

We analyzed the PPI Survey data at the specialty level rather than at the practice level. We took the following steps to analyze the PPI Survey data:

Created total compensation variables (Step 1). We created two total compensation variables at the specialty level. The first combined monetary compensation and benefits data for physicians and qualified health professionals (QHPs) provided in Table A, and the second combined monetary compensation and benefits data for administrative/clerical staff and clinical staff provided in Table B of the survey. We did this for cases with no missing values for total compensation in both tables.

Created flags for unusable data (Step 2). We created two flags to exclude unusable data. The first flag was for the number of reported physicians, and the second flag was for reported compensation and expenses.

The first flag looked at the number of physicians reported for each specialty. We generated a flag if the numbers of part-time and full-time physicians were both missing or equal to zero, or if one of the two was equal to 40, or unusually large. We flagged practices that entered 40 because this suggested that the respondent mistakenly read the question as number of hours worked rather than number of physicians in total. For those, we further investigated the other data reported to determine whether the 40 was likely an error.

The second flag examined how the practice reported compensation and expenses at the specialty level. If a practice specialty had missing or zero values (or a combination of missing and zero values) for all of the compensation and expense categories listed below, we flagged it as unusable data.

a. Administrative/clerical staff and clinical staff compensation (combined)
b. Physician compensation
c. Overhead expense
d. Medical supplies

e. Medical equipment
f. Non-billable drugs
g. Other expenses
h. Professional liability insurance
i. Information technology

We used these two flags to determine whether a specific observation (that is, data from a practice at the specialty level) could likely be used in the analysis. If all observations from a responding practice received a flag for the number of physicians reported (after resolving whether 40 was a mistake) or for not reporting information for any expense category, those observations were dropped from the analysis.

In total, 380 practices had at least one observation that we felt could likely be used in the analysis. There were 831 observations across the 380 practices. Together, these practices reported 18,086 physicians in Table A of the survey.

Generated total patient care hours provided in a year to serve as a denominator in the calculation of practice expense per hour worked for each specialty within a practice (Step 3). We generated estimates of total annual hours using the mean hours and weeks per year for part-time and full-time physicians multiplied by the numbers of part-time and full-time physicians, respectively, reported for a specialty by each practice.

Scaled down practice expense per hour worked for certain practices (Step 4). We scaled down practice expense per patient care hour for practices that did not allocate expenses to QHPs but did allocate QHP hours to each physician by specialty. We did this because in these cases, QHP expenses are integrated with physician specialty expenses and therefore overinflating physician practice expense. We scaled down practice expense for these practices based on the hours allocated to each physician specialty department for each type of QHP. For each physician specialty department, we summed (a) the total annual hours across all physicians in the specialty and (b) the total annual hours for QHPs that were allocated to the specialty. We then scaled down expenses by the ratio of a/(a+b). For example, if a specialty had 10,000 annual hours across all its physicians, and the practice allocated 5,000 annual QHP hours to this specialty, we multiplied the practice expense per hour of patient care provided by 10,000 / (10,000 + 5,000) = 0.67.

Expenses that we scaled down included administrative compensation, clinical compensation, information technology (IT), medical equipment, medical supplies, nonbillable drugs, overhead, and other expenses,

Trimmed outliers (Step 5). We trimmed outliers for expenses that were unreasonably low or high in the following categories: administrative compensation, clinical compensation, IT, liability, medical equipment, medical expenses, overhead expenses, other expenses, and provider compensation.

Imputed missing values (Step 6). We imputed missing values by expense category. Our approach to imputation varied by expense category to ensure that our imputation is as accurate as possible for each category. To determine the most appropriate approach for each expense category, we implemented regression analyses to investigate which practice and department characteristics are most significantly

related to each expense category within the same specialty. Here, we describe the imputation approach we implemented for each expense category:

- For administrative compensation, clinical compensation, medical equipment, medical supplies, non-billable drugs, overhead, IT, and other expenses, we imputed missing values based on the specialty and facility-billing specific mean.
- For liability, we imputed missing values based on the specialty-specific mean.
- For provider compensation, we imputed missing values based on practice size and the specialty-specific mean.

Created estimates of practice expense per hour of patient care provided by specialty (Step 7). We created estimates of practice expense per hour of patient care provided for each of the expense categories for 18 specialties. Due to small sample size for Radiation Oncology, we combined it with Radiology.

Lessons Learned

Through the process of sampling, fielding, and weighting the PPI Survey and the Physician Hours Survey, we learned several lessons that could improve future iterations of the survey. We have grouped these lessons into five thematic areas: evaluating the survey sampling unit, identifying practice contact information, engaging specialty societies in practice outreach, considering the survey length and complexity, and providing practices with more options for grouping specialties for the purpose of reporting expenses. We believe that addressing these lessons will improve the PPI Survey and result in a higher overall response rate.

Evaluating the survey sampling unit

We selected TINs as the practice sampling unit for several reasons. First, TINs that practices use to bill Medicare are available in MD-PPAS, as are other variables about practices that we used to select a representative sample, such as the percentage of claims billed within facility versus non-facility settings. Second, we thought practices would have financial information at the TIN level because practices use TINs for submitting taxes. Third, we have successfully used TINs to identify group practices in previous projects (Agency for Healthcare Research and Quality 2024).

Using TINs as the sampling unit led to two challenges: (1) practice organizations differ in terms of how they organize themselves into TINs and (2) practice organizations with multiple TINs cannot necessarily report expenses at the TIN-specialty level. With respect to the first challenge, some large multisite practice organizations have just one TIN for the entire organization, whereas others have multiple TINs corresponding to different practice sites or departments. The sample frame could thus include practice organizations with multiple TINs multiple times and therefore have multiple chances of being selected for the sample, whereas practices with only one TIN have only one chance. If a single practice organization had several TINs selected for the sample, it received several invitations to complete the survey, which was confusing at times. We attempted to identify all these cases and reach out to a practice representative so the representative was aware of why the practice received several invitations. However, it was not always possible to identify which practice sites were part of the same broad organization in cases when practice

sites did not share the same name. In addition, with respect to the second challenge, practice organizations use TINs for submitting tax information and billing insurers, but they cannot necessarily break out all expense information at the specialty level within each TIN. Several practice representatives told us it was easier for them to report expenses for the entire organization rather than for specific TINs within their organization.

If the AMA fields the PPI Survey again, it would be beneficial to consider alternative approaches to identifying a sampling unit so that each organization appears in the sample frame only one time. There are two potential options: (1) continuing to use the TINs but taking more time upfront to group TINs within the same organization together or (2) replacing TINs with another sampling unit. As noted, it is sometimes easy to identify the TINs affiliated with the same organization based on the TIN name. For example, in some cases all the TINs affiliated with the organization will have the organization's name (such as a name of a major health system) as part of the TIN name. However, that is not always the case. Sometimes if, for example, a larger organization acquires a practice that had its own TIN, that practice will maintain its own TINs and the TIN names are the names of the practices before the acquisition. In this case, it is very difficult to know that the practice is part of a larger organization without conducting additional online research to identify practice ownership information. The other option is to select a different sampling unit. One potential option is to use the IMS_ORG_ID in the IQVIA OneKey data set as the sampling unit. The disadvantage to using IMS_ORG_ID is IQVIA does not provide information regarding how this variable is constructed so we cannot be certain about whether, for example, there is only one IMS_ORG_ID for each large health system. Before deciding, additional work is necessary to assess the pros and cons of each option.

Identifying practice contact information

Although we were hopeful that the IQVIA OneKey data set would provide useful contact information for most practices in the sample, we discovered that the contact information in OneKey was often out of date, incorrect, or not available for the individuals within a practice most likely to have access to practice expense information. We are not aware of any alternative sources of practice contact information. If the AMA fields the PPI Survey again, it would be useful to confirm that there are no other sources of practice contact information. If there are alternative sources, before using that information the survey team should confirm that the contact information is accurate. If there are no other sources, the survey timeline needs to provide ample time and resources to support extensive locating efforts before fielding, including internet searches, outreach and communication with medical specialty societies, and phone calls to practices.

Engaging specialty societies in practice outreach

Before fielding the PPI Survey, the AMA reached out to specialty and state medical societies to obtain their support for the survey. As a result of this effort, more than 170 societies endorsed the survey. The invitation to participate in the survey noted the high volume of societies that endorsed the survey, mentioned the specific societies that are relevant to where the practice is located and the types of specialties the practice has, and provided an endorsement letter with the complete list of societies that supported the survey effort. Some societies also included information in their newsletter about the PPI Survey to encourage participation. We believe this was beneficial, but the overall response rate was still

low. To boost response, we received IRB approval to share lists of sampled practices with willing specialty societies through a secure transfer site so they could send targeted emails to the physicians from those practice that they have contact information for, asking them to speak with their practice finance or administrative staff about completing the survey. We believe this step was helpful in increasing the overall response rate and recommend repeating this step earlier in the fielding period.

Considering the survey length and complexity

During pre-testing, we heard that some large multispecialty practices spent up to 40 hours completing the PPI Survey because of the detailed expense reporting required at the specialty level. Small single-specialty practices spent considerably less time completing the survey, but the effort to collect the required expense information was still substantial and likely deterred practices from completing the survey. Overall, 1,523 practice respondents clicked on the survey link but never completed the screening questions and 317 practice respondents who completed the screener and were eligible for the survey never finished and submitted it. At least some percentage of these practices likely did not complete the survey because of the high burden associated with collecting all the relevant information.

If the AMA fields the PPI Survey again, it will be important to consider approaches to simplify the survey. and potentially collect relevant information through alternative means. With respect to collecting information through alternative means, the AMA could consider interviewing a few practices to understand if there are sources of information with expense data, such as tax returns, that practices would be willing to share confidentially to enable the AMA to calculate overall practice expense information that it would then share with the practice so respondents could confirm that they entered the data accurately.

Providing practices with specific options for grouping specialties for the purpose of reporting expenses

Prior to fielding the PPI Survey, the AMA recognized that it might be difficult for some practices to report all expense categories at the individual specialty level. As a result, we designed the PPI Survey to allow practices to report expenses for combinations of specialties, if needed. For example, if a practice could not break out expenses between family medicine and internal medicine physicians, the respondent could indicate that in the survey and then the survey would ask the respondent to report combined expenses for those two specialties in all subsequent tables. Overall, 64 of the 308 practices that submitted the survey reported combined expenses for at least 2 specialties. Additionally, of the 41 practices that were identified as multispecialty during sampling and submitted the survey, 21 reported combined expenses for at least 2 specialties. We also found that larger practices, meaning those with 50 or more physicians, were more likely than smaller practices to report combined expenses for at least 2 specialties.

In almost all cases, the specialties that practices combined for the purposes of reporting expenses were similar and aligned with the 19 collapsed specialty categories the AMA used for calculating practice expense per hours of patient care provided. However, in a handful of cases, practice respondents reported combined expenses for dissimilar specialties and, as a result, the AMA was unable to use the data they provided in the analysis.

Practice difficulty in reporting expenses for QHPs

Related, we also found that many practices were not able to separately report expenses for QHPs. For example, of the 380 practices whose data were used in the calculation of physician practice expense per hour, 145 reported that they had either nurse practitioners or physician assistants in their practice. However, only 39 of those practices were able to separately report non-provider compensation expenses for at least one of those QHP specialties. Among the 87 practices whose sampling specialty was in primary care, 29 reported nurse practitioners or physician assistants, and only 9 of them were able to separately report non-provider compensation expenses for them. Rather, practices with these two types of QHPs tend to embed their expenses with the physicians with whom they work.

References

Agency for Healthcare Research and Quality. "Compendium of U.S. Health Systems." Rockville, MD: AHRQ, September, 2017, revised February 2021. https://www.ahrq.gov/chsp/data-resources/compendium.html.

Agency for Healthcare Research and Quality, with D. Jones, K. Peckham, S. Nelson, R. Machta, and E. Rich. "Comparative Health System Performance Initiative: Compendium of U.S. Health Systems, 2018 Group Practice Linkage File, Technical Documentation." Rockville, MD: AHRQ, 2020; and Washington, DC: Mathematica, 2020. Available at https://www.ahrq.gov/sites/default/files/wysiwyg/chsp/compendium/2018-chsp-tin-linkage-file-tech-doc.pdf.

Agency for Healthcare Research and Quality. "Compendium of U.S. Health Systems." 2024. https://www.ahrq.gov/chsp/data-resources/compendium.html.

Baker, Lawrence C., M. Kate Bundorf, and Anne Beeson Royalty. "The Effects of Multispecialty Group Practice on Health Care Spending and Use." NBER Working Paper No. 25915. Cambridge, MA: National Bureau of Economic Research, 2019.

Furukawa, Michael F., Laura Kimmey, David J. Jones, Rachel M. Machta, Jing Guo, and Eugene C. Rich. "Consolidation of Providers into Health Systems Increased Substantially, 2016–18." *Health Affairs,* vol. 39, no. 8, 2020, pp. 1321–1325.

Kane, Carol. "Recent Changes in Physician Practice Arrangements: Shifts Away from Private Practice and Towards Larger Practice Size Continue Through 2022." Chicago, IL: AMA Policy Research Perspectives, 2023. Available at https://www.ama-assn.org/system/files/2022-prp-practice-arrangement.pdf.

Machta, Rachel M., James D. Reschovsky, David J. Jones, Laura Kimmey, Michael F. Furukawa, and Eugene C. Rich. "Health System Integration with Physician Specialties Varies Across Markets and System Types." *Health Services Research,* vol. 55, no. S3, 2020, pp. 1062–1072.

Mathematica Inc.

Our employee-owners work nationwide and around the world.

Find us at **mathematica.org** and **edi-global.com**.