ORIGINAL RESEARCH

The Results Are Only as Good as the Sample: Assessing Three National Physician Sampling Frames

Catherine M. DesRoches, Dr.P.H.[1], Kirsten A. Barrett, Dr.P.H.[1], Bonnie E. Harvey, B.A.[1], Rachel Kogan, M.P.H.[1], James D. Reschovsky, Ph.D.[1], Bruce E. Landon, M.D., M.B.A., M.Sc.[2], Lawrence P. Casalino, M.D., M.P.H., Ph.D.[3], Stephen M. Shortell, PhD, MPH, MBA[4], and Eugene C. Rich, MD[1]

[1]Mathematica Policy Research, Cambridge, MA, USA; [2]Department of Health Care Policy, Harvard Medical School, Boston, MA, USA; [3]Weill Cornell Medical College, New York, NY, USA; [4]Berkeley School of Public Health, University of California, Berkeley, CA, USA.

BACKGROUND: Databases of practicing physicians are important for studies that require sampling physicians or counting the physician population in a given area. However, little is known about how the three main sampling frames differ from each other.

OBJECTIVE: Our purpose was to compare the National Provider and Plan Enumeration System (NPPES), the American Medical Association Masterfile and the SK&A physician file.

METHODS: We randomly sampled 3000 physicians from the NPPES (500 in six specialties). We conducted two- and three-way comparisons across three databases to determine the extent to which they matched on address and specialty. In addition, we randomly selected 1200 physicians (200 per specialty) for telephone verification.

KEY RESULTS: One thousand, six hundred and fifty-five physicians (55 %) were found in all three data files. The SK&A data file had the highest rate of missing physicians when compared to the NPPES, and varied by specialty (50 % in radiology vs. 28 % in cardiology). NPPES and SK&A had the highest rates of matching mailing address information, while the AMA Masterfile had low rates compared with the NPPES. We were able to confirm 65 % of physicians' address information by phone. The NPPES and SK&A had similar rates of correct address information in phone verification (72–94 % and 79–92 %, respectively, across specialties), while the AMA Masterfile had significantly lower rates of correct address information across all specialties (32–54 % across specialties).

CONCLUSIONS: None of the data files in this study were perfect; the fact that we were unable to reach one-third of our telephone verification sample is troubling. However, the study offers some encouragement for researchers conducting physician surveys. The NPPES and to a lesser extent, the SK&A file, appear to provide reasonably accurate, up-to-date address information for physicians billing public and provider insurers.

BACKGROUND

Obtaining accurate physician contact information is increasingly important as researchers attempt to improve our understanding of the extent of change in the U.S. health care system, distribution of physicians, and availability and access to resources. Given physicians' role in the health care delivery system, understanding their experiences, attitudes, and behaviors is enormously important for evaluating the success of these public and private efforts.[1] However, the reliability of existing sampling frames can present a challenge to ensuring representativeness and decreasing bias in physician surveys.

Mailed surveys offer a cost-efficient method of collecting a significant amount of data from a large number of individual physicians about their clinical practice.[2] However, surveying physicians presents a number of methodological challenges. Foremost is the identification of the physician sample frame (that is, a list of physicians from which a sample is drawn). Ideally, any survey sample should be drawn from a frame that contains the universe of potential respondents and up-to-date and accurate contact information.[3] In practice, physician sampling frames are highly variable and can suffer from under coverage, meaning the frame does not contain the universe of physicians, or over coverage, meaning the frame contains physicians who are no longer in practice, such as those who are retired or deceased.[4] In addition to coverage issues, sample frames may include erroneous and/or duplicate address and telephone information, further reducing their utility.

Procedures for updating and maintaining physician sampling frames are variable. Physicians are entered into the National Provider and Plan Enumeration System (NPPES) when they receive a National Provider Identifier (NPI) number (required for billing public and private insurers). Since the address in the NPPES is used for billing, physician organizations have an incentive to update address information. Further, individual physicians are required by the Centers for Medicare & Medicaid Services (CMS) to update their contact information within 30 days of an address change, but this is not uniformly enforced.[5]

The American Medical Association (AMA) captures extensive information about physicians when they enter medical school, and includes professional certification information

gathered from State licensing boards as they enter practice. Further, the AMA makes efforts to verify and update physician contact information; however, physicians are not required to update their contact information.[6,7]

SK&A's database contains information on physicians in office-based practices, including some practices that are owned by or located in hospitals. Unlike the NPPES and AMA files, SK&A attempts to verify contact information for all physician practices via phone every 6 months. Physicians within practices are enumerated during the call. In addition, SK&A provides practice-level variables that are not available in either the NPPES or AMA files (patient volume, number of providers, site specialty, and ownership), although we do not attempt to verify the accuracy of the practice-level information in this article.[8]

The coverage-related issues inherent in physician sampling have been documented; and research has discussed some of the limitations of existing sample frames.[4,9–11] However, relatively little is known about how specific sampling frames differ from each other in terms of the coverage and accuracy of contact information. In this study, we compared the NPPES file, the AMA Masterfile, and the SK&A physician file to determine how each fared with regard to the coverage and accuracy of key variables, including address and specialty information.

DESIGN

The characteristics of the three sampling frames used in this study are outlined in Table 1. We developed our analytic file
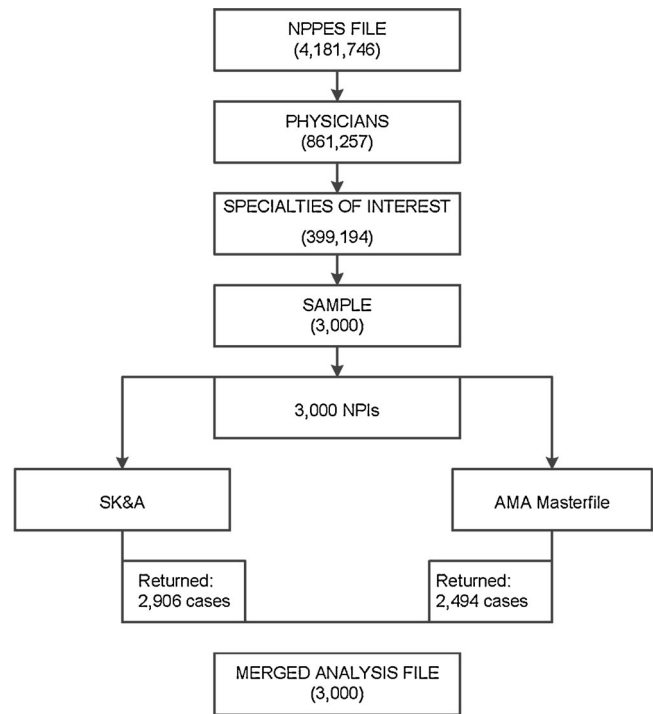

Figure 1 Building the analytic file.

using a four-step process (Fig. 1). First, we downloaded the NPPES file, current as of December 2013. The file contained approximately 4.1 million records, of which 861,257 were physicians. Second, we limited the file to physicians reporting a primary specialty of family medicine, internal medicine,

Table 1. Data File Characteristics

Database	American Medical Association masterfile	National Plan and Provider Enumeration System	SK&A
Coverage	1 million physicians, residents, medical students + 66,000 osteopaths	4.1 million physicians and nonphysicians; 861,257 physicians	740,000 office-based physicians
Cost	Variable based on the number of records and variables purchased	Free	Variable based on the number of records and variables purchased
Individual-level variables	Name, address, DOB, birthplace, medical school, specialty, board certifications, employer, professional medical activities, principal hospital and group affiliations[a]	National Provider Identifier (NPI); provider or business name, provider mailing address, health care provider taxonomy code, provider enumeration date, NPI deactivation reason code, NPI deactivation date, gender, provider license number, State, and authorized official contact information[b]	Provider name, provider address, specialty, patient volume
Practice-level variables	Type of practice, group practice locations	Provider business location address	Practice size, HIT adoption, ACO participation
Advantages	Provides some variables for stratification; AMA makes attempts to verify and upate physician information; one of the most common sample frames used for physician surveys.	Free; information is publicly available; includes all physicians who request reimbursement through Medicare or private insurance; required updating	Phone verification of cases every 6 months; has best information on practice-level variables
Disadvantages	Expensive; limited ability to prestratify, conduct nonresponse weighting and oversampling	Does not include physicians who do not bill third-party payers. Few variables for stratification available. Up dating is required but not enforced	Expensive; only includes office-based physicians, excludes hospitalists and physicians in other nontraditional settings (such as telemedicine). Not all data are phone-verified

[a] *American Medical Association. AMA Physician Masterfile. Available at* http://www.ama-assn.org/ama/pub/about-ama/physician-data-resources/physician-masterfile.page
[b] *Centers for Medicare & Medicaid Services. National Provider Identifier Standard (NPI). Available at:* http://www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/NationalProvIdentStand/index.html?redirect=/nationalprovidentstand *(Accessed 15 April 2015)*

pediatrics, orthopedic surgery, radiology, or cardiology. These specialties were chosen either because they (1) have implications for the primary care workforce, (2) account for a substantial proportion of Medicare spending, and/or (3) represent physicians who provide care in a variety of settings and types of practice organizations.[10,12] This process resulted in a file containing 399,194 physicians. Third, we randomly selected 500 cases from each of the six specialties for a total sample size of 3000 physicians. Fourth, we sent the sampled NPIs to SK&A and Medical Marketing Services (MMS), the vendor for the AMA Masterfile. We asked them to match NPIs to physicians in their data file and return the file to us with the matched physician, address and telephone. Contact information provided by SK&A and MMS was merged with that in the NPPES file for analysis.

Address Verification Via Phone Calls

After creating the merged sample file, we randomly selected 200 physicians from each specialty for telephone verification ($N=1200$). We called the listed telephone numbers and attempted to confirm that the physician could be contacted via mail at the address contained within the data file(s). In cases where the sample frames each had a different practice location name and address, we called each location a maximum of three times during normal business hours to confirm the contact information and to account for the possibility that a physician might practice at more than one location. When the phone information contained in the data file was incorrect (that is, out of service, a fax line, or a wrong number), we conducted internet searches to find alternative phone numbers for physicians.

Analysis

There is no gold standard to compare against to determine the accuracy of physician sample frame address information. For this study, we compared the information in SK&A and the AMA Masterfile to the NPPES for consistency and verified the contact information for a subsample of physicians. We conducted two-way and three-way comparisons on address fields and specialty fields. We compared the address information across the three data files using a combination of geocoding and manual inspection. Geocoding creates markers for the longitude and latitude of a given address (street address, city, state, and zip code) and provides a more efficient method of location matching than matching on alphanumeric characters. Further, we manually inspected addresses to ensure that matching addresses were not missed during geocoding. Addresses were coded as matching across databases if they had the same geocoded markers for latitude and longitude, or if the listed addresses were matched on street address, city, and zip code.

In the analyses, cases were included only if they matched on NPI across all three files for the three-way comparison, and in both files for the two-way comparisons. In addition, SK&A cases that were not contained within the company's phone verified practice-level database, and therefore had never been verified, ($n=1085$) were excluded from these analyses across data files. Their inclusion would have resulted in a misleading rate of matching information between the NPPES and SK&A files, as many of the non-phone-verified cases in the SK&A file have data fields populated with NPPES data. We next examined the extent to which specialty designations in the AMA and SK&A files match those in the NPPES file.

Finally, we merged the files together sequentially to determine the contribution of each to the overall sample, to determine if a combined sample frame would offer a more robust file than a single frame. We started with the NPPES and then added cases unique to the SK&A (using only the cases phone verified by the company), and then those unique to the AMA.

RESULTS

NPI Matching

SK&A returned a file with 2906 (97 %) cases that matched the NPPES file on NPI; of these, 1821 (63 %) had been phone verified by SK&A. In the remaining 1085 (47 %) cases, contact information was derived from government records (NPPES, State licensing boards, DEA). In subsequent analyses, we limited the SK&A file to the 1821 cases that were phone verified by the vendor. MMS returned a file with 2494 (83 %) matching cases from the AMA Masterfile that matched the NPPES file; 506 (17 %) NPIs provided to MMS were not found in the AMA Masterfile.

Fifty-five percent of physicians ($n=1655$) were contained in all three data files. Rates of matching on NPI did not vary significantly by specialty among physicians that were contained in both the NPPES in the AMA data files (86 % for of pediatricians matched on NPI to 82 % of family medicine physicians, radiologists, and cardiologists). (See Technical Appendix for full results). However, rates of matching did vary by specialty in the SK&A data when compared to the NPPES, from a low of 50 % matching on NPI for radiologists to a high of 72 % for cardiology (See Technical Appendix for full results).

Address Matching

As a complete address block (address, city, state, and zip code), 33 % of cases matched on all four elements across all three databases. We next conducted a series of two-way comparisons, as shown in Table 2, comparing the NPPES to the AMA Masterfile and the SK&A database. First and last names matched at a very high rate across all two-way comparisons, as

Table 2. Percent of Physicians Matching on Name, Address, and Specialty Designation

	AMA vs. NPPES $n=2494$	NPPES vs. SK&A[a] $n=1821$
	% (N)	% (N)
Contact information		
Last name	98 % (2457)	100 % (1819)
First name	100 % (2486)	100 % (1815)
Street address	33 % (817)	70 % (1284)
City	58 % (1460)	81 % (1481)
State	91 % (2268)	97 % (1773)
Zip	44 % (1094)	77 % (1409)

[a]SK&A phone-verified database

did State. However, rates of matching were lower for other address fields and varied across two-way comparisons. When the AMA was compared to NPPES, we found a matching street address for approximately one-third of our sampled physicians (33 %). When NPPES was compared to the phone verified SK&A data, the matching rate on street address improved to 70 %. We found a similar pattern for city and zip code, with the NPPES versus SK&A comparison resulting in a higher percentage of matches than the other two-way comparisons.

Specialty Matching

Specialty information in the sample frame is often used for sample stratification. Therefore, we next examined the extent to which specialty designation varied across the data files. That is, we assessed the extent to which a physician designated as a cardiologist in the NPPES file had the same designation in SK&A and AMA Masterfile (Fig. 2). Specialty designation was the same for 78 % of the physicians contained in all three files, regardless of address matching. A series of two-way comparisons showed that the specialty match between NPPES and AMA and NPPES and SK&A was 80 and 91 %, respectively. In the NPPES versus AMA comparison, the lowest match rate was for radiology and the highest was for family medicine. For the NPPES versus SK&A comparison, internal medicine and radiology accounted for the lowest and highest match rates.

Address Verification Via Phone Calls

In our attempt to verify physician address information, we placed calls to all cases in the subsample of 1200 physicians from our total sample of 3000; 65 % of the 1200 calls made ($n=777$) resulted in confirmation of street, city, state, and zip code information contained in at least one of the data files. Overall, SK&A (85 %) and the NPPES (86 %) had the highest rates of correct address information among our confirmed cases, whereas the AMA Masterfile included correct contact information for less than half of the physicians we were able to confirm (42 %). In the NPPES, the proportion of physicians by specialty with correct information was highest in general internal medicine (94 %) and lowest for radiology (72 %). For SK&A, percentages were highest for cardiology (92 %) and family medicine (91 %), and lowest for radiology (80 %) internal medicine (79 %). In the AMA Masterfile, confirmation of contact information was lowest in radiology (32 %) and highest in family medicine (54 %). (See Technical Appendix Table 8 for full results.)

We were unable to confirm the remaining 432 physicians. Cases were most often unconfirmed because the telephone was not answered during any of our call attempts (25 %). Other reasons for coding a physician as unconfirmed included: physician had left the practice or was not at the location (21 %), non-working number (19 %), or the listed number was a switchboard or answering service that was unable to confirm the physician contact information (9 %). As shown in Table 3, the percentage of unconfirmed physicians varied by specialty from 49 % in internal medicine to 25 % in family medicine.

Combining Data Files

We examined the marginal benefit of adding the SK&A and AMA files to the NPPES file in terms of the number of confirmed cases each additional file added. Of the 777 confirmed cases, 671 were correct in the NPPES database (86 %).
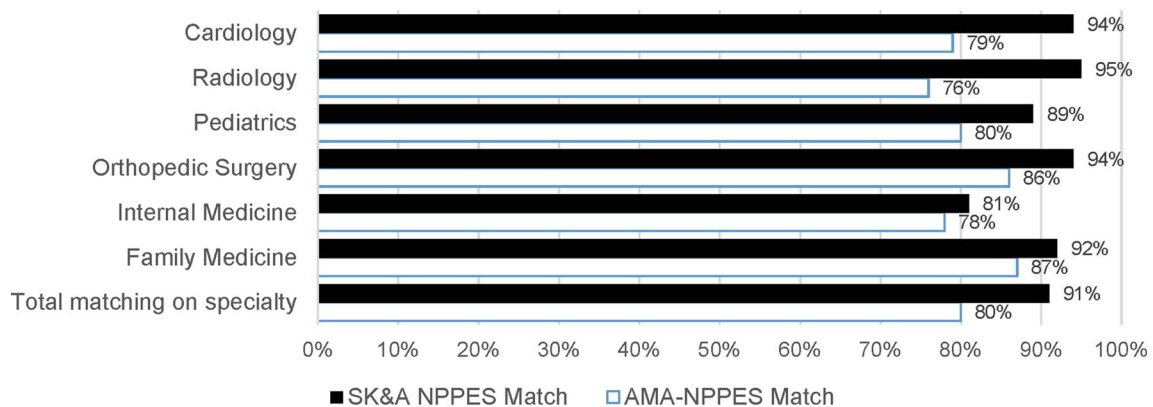

Figure 2. Percent of physicians matching on specialty designation.

Table 3 Percent of Unconfirmed Physicians by Specialty

	Total	Physician with unconfirmed contact information	
		N	*%*
Total	1,200	423	35 %
Family medicine	200	48	24 %
Internal medicine	200	98	49 %
Orthopedic surgery	200	62	31 %
Pediatrics	200	56	28 %
Radiology	200	85	43 %
Cardiology	200	74	37 %

Adding the SK&A file to the NPPES file resulted in an additional 85 correct addresses, or 11 % of our confirmed cases. Combining NPPES and AMA files together resulted in 46 additional cases, or 6 % of our confirmed cases. The SK&A and AMA files together yielded 106 cases that would have been excluded if the NPPES were used as a sample frame in isolation (14 %).

DISCUSSION

Most large-scale, high-response-rate physician survey efforts such as the National Ambulatory Medical Care survey and the Physician Component of the Community Tracking Survey begin with the AMA Masterfile.[12,13] Findings from our research suggest that researchers should consider using other sampling frames, as the AMA Masterfile had the lowest contact information accuracy of the three sources we examined and the highest percentage of unconfirmed physicians. This suggests that a majority of the physicians either listed a location that we were unable to verify (potentially a nonpractice location or home address), or failed to update their contact information in the AMA Masterfile when they changed locations or retired. It is possible that the physicians we were unable to confirm in the AMA Masterfile differ systematically from those we were able to reach, resulting in biased results when the Masterfile is used as a sampling frame; however, our study was not designed to investigate this potential bias.

Overall, SK&A and the NPPES had the highest rates of correct contact information among our telephone-confirmed cases. However, this varied by specialty. Since the SK&A database focuses on office-based physicians, physicians who commonly work in hospital-based specialties—for example, radiologists—are underrepresented. Our findings suggest that the contact information contained within the NPPES file may be more accurate than that found in either the SK&A file or the AMA Masterfile. And, as evidenced by the number of cases matching between AMA and the NPPES and SK&A and the NPPES, the coverage is best with the NPPES data file.

Adding costly and proprietary SK&A and AMA data to the NPPES (which is publicly available) file resulted in a small increase (14 %) in the overall number of cases with correct contact information, based on our telephone verification. However, the NPPES lacks information needed for sample stratification other than by medical specialty, such as professional age (contained in the AMA Masterfile) and practice setting characteristics (contained in the SK&A data file), although in the case of the latter, we cannot verify the accuracy of this information. Still, the accuracy of the NPPES contact information and the savings that researchers could achieve by using a free data source may be worth the tradeoff. Further research should explore the utility of linking organizational-level variables such as size, ownership, and use of electronic health records available in the SK&A data file to individual physician records drawn from the NPPES.

Several limitations to our study should be considered. First, we included only physicians in six specialties; therefore, we cannot generalize to all physicians. Second, the NPPES only includes physicians who bill third-party payers, although relatively few physicians practicing in the U.S. fail to accept any insurance.[14] Third, the AMA Masterfile allows physicians to list their preferred mailing address. The use of a preferred mailing address may have affected both the rate of matching to the NPPES and our ability to confirm contact information if physicians use an address other than the practice location. The AMA does offer information on primary office location, however, to our knowledge this information is not widely used by survey researchers. Future work on the completeness and accuracy of the primary office location information is necessary to assess its usefulness for survey researchers. Finally, there is no gold standard to compare the accuracy of the address and telephone number information in each of the three data files. We chose to compare the AMA and SK&A files to the NPPES, but we cannot rule out the possibility that the information in the NPPES is incorrect for physicians we were unable to confirm by telephone.

Our findings indicate that none of the data files in this study are perfect; the fact that we were unable to reach one-third of our telephone verification sample is troubling, as was the variation in confirmation by specialty. Our inability to contact physicians listed in these databases may be due to a number of reasons, including a failure to update contact information (AMA Masterfile and NPPES). In addition, we may have been unable confirm physicians contact information due to "over coverage," meaning that the database includes physicians who are retired or deceased. Finally, data on physicians may be missing from the SK&A phone verified database simply because SK&A has not included the practice in the phone verification process yet. In order to assess the potential for biased results, further research is needed to determine if these missing physicians differ systematically in some way from those we were able to locate. However, the study findings offer encouragement for researchers conducting physician surveys. The NPPES currently appears to provide accurate, up-to-date

contact information for physicians billing public and private payers. Using this free data file would allow researchers to reallocate financial resources from purchasing sample to other aspects of survey administration that could positively impact response rates. Given physicians' important role in current efforts to transform the health care delivery system, obtaining accurate, generalizable information about their attitudes, behaviors, and experiences, as well as the characteristics of the organizations in which they practice, will continue to be critically important.

TECHNICAL APPENDIX

Table 4 Percentage of Cases Matching on NPI by Specialty

	NPPES	AMA	SK&A[a]
		Cases matching the NPPES on NPI	Cases matching the NPPES on NPI
	N	% (N)	% (N)
All	3000	83 % (2494)	61 % (1821)
Family Medicine	500	82 % (409)	62 % (311)
Internal Medicine	500	84 % (422)	52 % (258)
Orthopedic Surgery	500	83 % (413)	71 % (353)
Pediatrics	500	86 % (430)	58 % (288)
Radiology	500	82 % (408)	50 % (252)
Cardiology	500	82 % (412)	72 % (359)

[a]*SK&A phone-verified database*

Table 5 Percentage of NPPES Physicians Found in Both the AMA Masterfile and SK&A Data File

N=1655	AMA vs. NPPES vs. SKA[a]
	NPI matching
Contact information	% (N)
Last name	99 % (1638)
First name	99 % (1646)
Street address	33 % (561)
City	58 % (965)
State	95 % (1572)
Zip	45 % (747)
Specialty	78 % (1290)

[a]*SK&A phone-verified database*

Table 6 Two-Way Matching on Elements of Contact Information

Information	AMA vs. NPPES n=2494 NPI matching % (N)	AMA vs. SK&A[a] n=1655 NPI matching % (N)	NPPES vs. SK&A[a] n=1821 NPI matching % (N)
Last name	98 % (2457)	99 % (1640)	100 % (1819)
First name	100 % (2486)	99 % (1646)	100 % (1815)
Address	33 % (817)	38 % (636)	70 % (1284)
City	58 % (1460)	63 % (1048)	81 % (1481)
State	91 % (2268)	97 % (1602)	97 % (1773)
Zip	44 % (1094)	49 % (816)	77 % (1409)
Specialty	80 % (1992)	82 % (1360)	91 % (1659)

[a]*SK&A phone-verified database*

REFERENCES

1. **Berenson RA, Kaye DR.** Grading a physician's value—the misapplication of performance measurement. N Engl J Med. 2013;369(22):2079–2081.
2. **Aday LA.** Designing and conducting health surveys. San Francisco: Jossey-Bass Publishers; 1991.
3. **Kish L.** Survey sampling. New York: Willey; 1995.
4. **DiGaetano R.** Sample frame and related sample design issues for surveys of physicians and physician practices. Eval Health Prof. 2013;36(3):296–329.
5. Centers for Medicare & Medicaid Services. 2014 National Plan and Provider Enumeration System. https://nppes.cms.hhs.gov/NPPES/Welcome.do. Accessed April 15, 2015.
6. American Medical Association. 2014 American Medical Association Masterfile. http://www.ama-assn.org/ama/pub/about-ama/physician-data-resources/physician-masterfile.page. Accessed April 15, 2015.
7. American Medical Association. http://www.mmslists.com/news-articles/article.asp?ID=98. Accessed April 15, 2015.
8. **Dunn A, Shapiro AH.** Do physicians possess market power? J Law Econ. 2014;57(1):159–193.
9. **Clark J, Digaetano R.** Using the national provider identification file as the sampling frame for a physician survey. Joint Statistical Meeting; August 2014; Boston. Alexandria, VA: American Statistical Society; 2014.
10. **Klabunde CN, Willis GB, McLeod CC, et al.** Improving the quality of surveys of physicians and medical groups: a research agenda. Eval Health Prof. 2012;35(4):477–506.
11. **McLeod CC, Klanbunde CN, Willis GB, Stark DB.** Health care provider surveys in the United States, 2000–2010: a review. Eval Health Prof. 2013;36(1):106–126.
12. Centers for Disease Control and Prevention. 2010 NAMCS scope and sample design. www.cdc.gov/nchs/ahcd/ahcd_scope.htm#namcs_scope. Accessed April 15, 2015.
13. **Strouse R, Potter F, Davis T, et al.** HSC 2008 Health Tracking Physician Survey Methodology Report, Technical Publication, vol. 77. Washington, D.C: Center for Studying Health System Change; 2009.
14. Medicare Payment Commission. Report to Congress: Medicare Payment Policy. Chapter 4: Physicians and other health professional services: Assessing payment adequacy and updating payments. March 2014. Available at: http://www.medpac.org/documents/reports/mar14_ch04.pdf?sfvrsn=0. Accessed November 20, 2014. Accessed April 15, 2015.

Table 7 Percentage of Physicians With Matching Specialty Designation Across Data Files

	NPPES-AMA matching cases	NPPES-SK&A[a] matching cases		
	Match on NPI	Match on specialty	Match on NPI	Match on specialty
	N	% (*N*)	*N*	% (*N*)
Total	2494	80 % (1992)	1821	91 % (1659)
Family Medicine	409	87 % (357)	311	92 % (285)
Internal Medicine	422	78 % (331)	258	81 % (210)
Orthopedic Surgery	413	86 % (357)	353	94 % (331)
Pediatrics	430	80 % (346)	288	89 % (256)
Radiology	408	67 % (275)	252	95 % (239)
Cardiology	412	79 % (326)	359	94 % (338)

[a]*SK&A phone-verified database*

Table 8 Percentage of Confirmed Physicians with Correct contact Information by Specialty and Data File

	Number of confirmed physicians	Percent of confirmed physicians with correct contact information in each data file		
		AMA masterfile	NPPES	SK&A
Family Medicine	152	82 (54 %)	134 (88 %)	140 (92 %)
Internal Medicine	102	38 (37 %)	96 (94 %)	81 (79 %)
Orthopedic Surgery	138	70 (51 %)	87 (85 %)	116 (84 %)
Pediatrics	144	50 (35 %)	131 (91 %)	120 (83 %)
Radiology	115	37 (32 %)	83 (72 %)	92 (80 %)
Cardiology	133	56 (42 %)	114 (86 %)	121 (91 %)