Creating a National Provider Identifier (NPI) to Unique Physician Identification Number (UPIN) Crosswalk for Medicare Data

Helen M. Parsons, PhD, MPH[1] **[Assistant Professor]**,

Department of Epidemiology and Biostatistics, School of Medicine, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, Mail Code 7933San Antonio, TX 78229-3900, Phone: (210) 567-0854, Fax: (210) 567-0921

Lindsey R. Enewold, PhD, MPH[2],

Healthcare Delivery Research Program, National Cancer Institute, 9609 Medical Center Drive, Room 3E428, Bethesda, MD 20892-9704, Phone: (240) 276-6856

Robert Banks[3],

Information Management Services, Inc., 3901 Calverton Blvd, Suite 200, Calverton, MD 20705, Phone: (301) 680-9770

Michael J. Barrett, M.S.[3], and

Information Management Services, Inc., 3901 Calverton Blvd, Suite 200, Calverton, MD 20705, Phone: (301) 680-9770

Joan L. Warren, Ph.D.[2]

Healthcare Delivery Research Program, National Cancer Institute, 9609 Medical Center Drive, Room 3E428, Bethesda, MD 20892-9704, Phone: (240) 276-6815

Abstract

Introduction—Many health services researchers are interested in assessing long-term, individual physician treatment patterns, particularly for cancer care. In 2007, Medicare changed the physician identifier used on billed services from the Unique Physician Identification Number (UPIN) to the National Provider Identifier (NPI), precluding the ability to use Medicare claims data to evaluate individual physician treatment patterns across this transition period.

Methods—Using the 2007-2008 carrier (physician) claims from the linked Surveillance, Epidemiology and End Results (SEER) cancer registry-Medicare data and Medicare's NPI and UPIN Directories, we created a crosswalk that paired physician NPIs included in SEER-Medicare data with UPINs. We evaluated the ability to identify an NPI-UPIN match by physician gender and specialty.

Corresponding Author: Helen M. Parsons, PhD, MPH, Assistant Professor, Department of Epidemiology and Biostatistics, School of Medicine, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, Mail Code 7933San Antonio, TX 78229-3900, Phone: (210) 567-0854, Fax: (210) 567-0921, parsonsh@uthscsa.edu.
[1]The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, Mail Code 7933, San Antonio, TX 78229-3900
[2]Healthcare Delivery Research Program, National Cancer Institute, 9609 Medical Center Drive, Room 3E428, Bethesda, MD 20892-9704
[3]Information Management Services, Inc., 3901 Calverton Blvd, Suite 200, Calverton, MD 20705

Conflicts of Interest: The authors declare no conflicts of interest.

Results—We identified 470,313 unique NPIs in the 2007-2008 SEER-Medicare carrier claims and found a UPIN match for 90.1% of these NPIs (n=423,842) based on three approaches: 1) NPI and UPIN co-reported on the SEER-Medicare claims, 2) UPINs reported on the NPI directory or 3) a name match between the NPI and UPIN directories. 46.6% (n=219,315) of NPIs matched to the same UPIN across all three approaches, 34.1% (n=160,277) agreed across two approaches, and 9.4% (n=44,250) had a match identified via one approach only. NPIs were paired to UPINs less frequently for women and primary care physicians compared with other specialists.

Discussion—NCI has created a crosswalk resource available to researchers that links NPIs and UPINs based on the SEER-Medicare data. Additionally, the documented process could be used to create other NPI-UPIN crosswalks using data beyond SEER-Medicare.

Keywords

Introduction

Many health services researchers are interested in determining the practice and treatment patterns of physicians over time.[1-6] Interest in studying these temporal patterns will likely increase, as more research is directed toward understanding the impact of new health care payment models. The Medicare data are an important resource to evaluate these types of questions as they provide healthcare claims from physicians and treating facilities for the more than 50 million current Medicare beneficiaries. These claims identify both the performing and referring physician for each service.[7-9] Prior to the introduction of the Health Insurance Portability and Accountability Act (HIPAA),[10] Medicare used a Unique Physician Identification Number (UPIN) to identify the physicians and other healthcare providers who billed Medicare for services provided to beneficiaries.[11,12] Each physician, as an individual, should have had only one UPIN throughout their career, regardless of the locations or states in which they practiced.[10,11] Beginning in 2007, the National Provider Identifiers (NPIs) replaced UPINs as the required identifier for physicians submitting bills to Medicare.[11,12] The Center for Medicare and Medicaid Services (CMS) began issuing NPIs in October of 2006 and required the use of only NPIs for Medicare provider reimbursement beginning after May 2007. As with the UPIN, a physician's NPI should not change once issued and remains constant regardless of job or location changes.[11,12]

The implementation of a new provider identification system introduced challenges for researchers who are interested in evaluating individual physician treatment patterns that span the transition from UPIN to NPI. Without a method to track individual physicians, researchers cannot assess trends in physician treatment patterns over time. Tracking the same physicians over time is particularly important for researchers interested in cancer treatment and survivorship. Cancer patients often require treatment from a range of specialists over extended periods of time.[13-15]

Physician practice patterns for cancer patients have been frequently studied using the Surveillance, Epidemiology and End Results (SEER)-Medicare data, a population based resource that includes linkage of cancer registry data for Medicare beneficiaries with cancer

to their Medicare claims.[8,16-18] The SEER-Medicare data, available back to 1991, are released to researchers for cancer health services research. Currently, there over 1,250 peer-reviewed publications using these data.[19-38] To date, SEER-Medicare studies that have focused on individual physician treatment patterns have been temporally limited because a crosswalk that links individual physician UPINs and NPIs was not available.

To provide the ability to track individual physician treatment patterns over time for cancer research, we developed a SEER-Medicare physician NPI-UPIN crosswalk file using the National Claims History (NCH) files from the SEER-Medicare data and the CMS NPI and UPIN Directories. We then evaluated the ability to identify an NPI-UPIN match by physician gender and specialty.

Methods

Data

Data for this project came from the linked SEER-Medicare data and the CMS NPI and UPIN Directories. The SEER data come from National Cancer Institute (NCI) sponsored cancer registries that collect cancer incidence, demographic and cause of death information for persons with cancer. The SEER registries currently collect data from population-based cancer registries covering approximately 26% of the US population.[39] The Medicare claims include all covered health care services from the time of a person's Medicare eligibility until death. Medicare receives claims from all approved providers. Persons in the SEER data are linked biennially to the Medicare data. Over 93% of persons in SEER who are at least age 65 years are matched to Medicare enrollment data. The SEER-Medicare data are held by NCI and are released to researchers, following the data use requirements of CMS and the SEER registries. The SEER registries require that prior to release, the physician UPINs and NPIs on the Medicare claims are encrypted in a manner that is unique to each physician but prevents the physician from being re-identified.

For our study, we used Medicare Part B carrier claims. The carrier files within the Medicare data include primarily physician bills for services provided to Medicare beneficiaries, although the files also include claims from other non-institutional providers of Medicare services including physician assistants, clinical social workers, nurse practitioners, independent clinical laboratories and stand-alone ambulatory surgical centers.[8] In developing our crosswalk, we included only identifiers for physicians, as they direct the majority of treatment decisions in oncology. Further, prior to the introduction of NPIs, other providers (i.e., nurse practitioners) commonly billed for services under a group rather than individual identifier, making the link between NPIs and UPINs on an individual level challenging. We included claims from 2007-2008 as, during this time of transition, each claim could include the NPI and the UPIN for the same physician.

In addition to Medicare claims, we used information from the NPI Directory (National Plan and Provider Enumeration System (NPPES)) and the UPIN Directory, which are maintained by CMS. The NPI Directory includes a listing of all registered NPIs and their associated information including the physician's name, gender, specialty, practice location, state of license and other identifiers at the time of application.[6] While not required by CMS, the

application for an NPI does allow the physician to list voluntarily any previous UPIN used, which is then included on the NPI Directory. The NPI Directory also includes taxonomy codes that identify the type of provider (physician, nurse practitioner, etc). The UPIN Directory contains similar information for physicians who are enrolled in the Medicare Program, including full name, specialty, and practice location.

Creating the NPI-UPIN Crosswalk File

The creation of the NPI-UPIN crosswalk file was intended to assist SEER-Medicare researchers track individual physicians over time; therefore, we first identified all NPIs included on the 2007-2008 SEER-Medicare carrier claims. Only paid claims were included to limit the inclusion of erroneous claims. We, then, linked the list of identified NPIs from the claims with the CMS NPI Directory and retained only the NPIs that were associated with a taxonomy code indicating that the provider was a physician (i.e. '20XXXXXXX').[7] Thereby, we excluded group NPIs and non-physician practitioners (e.g., nurse practitioners, physician assistants). Following identification of this unique set of physician NPIs, we used three independent approaches to identify a UPIN for each physician NPI, as described below and outlined in Figure 1.

Approach 1: Matching an NPI to a UPIN in the SEER-Medicare Claims—For our first approach, we used paid claims from the 2007-2008 SEER-Medicare carrier files, as providers could list a UPIN, NPI or both as the physician identifier for billing during this timeframe (Figure 1). We then kept only those claims that had both an NPI and a UPIN reported on the same claim. We determined that an NPI was matched to a UPIN only if the same NPI and UPIN combination was reported on all of these billed claims. We did not consider an NPI matched to a UPIN in situations where there was not a one to one NPI-UPIN match (e.g., one NPI matched to a different UPIN across claims).

Approach 2: Identifying UPINs reported on the NPI Directory—For Approach 2, we referenced the CMS NPI Directory to determine if physicians with NPIs from the 2007-2008 SEER-Medicare carrier claims had listed a previous UPIN when applying for an NPI. We determined that an NPI was matched to a UPIN within this approach if one, and only one, UPIN was reported on a physician's NPI Directory listing.

Approach 3: Used the NPI and UPIN Directories—For our final approach, we examined whether we could match UPINs to our list of NPIs based on name matching techniques. We abstracted the first name, middle initial and last name for the previously identified physician NPIs in the NPI Directory and separately abstracted the first name, middle initial, and last name for UPINs in the UPIN Directory. UPINs were then linked to the NPIs by last name, first name, and middle initial, if reported. The matching was done using only the alphabetic portion of the names, compressing blanks, digits and special characters. Middle initial was ignored if it was missing in either of the directories. NPI and UPIN names were considered the same person if there was an exact name match between the two directories. If the same name was associated with more than one NPI or UPIN (e.g., more than one Jane A. Smith), we determined that an NPI-UPIN match could not be made using this approach.

Final Crosswalk—We used the three above approaches to identify corresponding UPINs for our unique set of physician NPIs. We then assessed across-approach agreement (was the same UPIN assigned to the NPI across approaches). For each NPI, we assigned an assessment of how confident we were that the accurate matching UPIN was identified. If an NPI was matched to the same UPIN via all three approaches, we had very strong confidence that the NPI-UPIN match was accurate. If an NPI was matched to the same UPIN via two approaches but the third approach identified a different UPIN, we had strong confidence in the corroborated NPI-UPIN match. If an NPI was matched to the same UPIN via two approaches and the third approach failed to identify a corresponding UPIN, then we also had strong confidence in the identified NPI-UPIN match. If only one approach successfully identified a UPIN, then we had moderate confidence in the NPI-UPIN match. If, upon across-approach comparison, an NPI was associated with any unresolvable conflicts (i.e., >1 NPI matched to 1 UPIN or 1 NPI matched to >1 UPIN) then no UPIN was ultimately assigned. The crosswalk, therefore, does not include any duplicate NPIs or UPINs. The final NPI-UPIN crosswalk file includes all unique NPIs; the corresponding UPIN, if successfully matched; and an assessment of match confidence (very strong, strong, moderate, unresolvable conflict, or no match).

Statistical Analysis

We used chi-square analysis to evaluate whether match confidence differed by physician gender (male vs. female) and specialty according to information reported in the NPI Directory. Physician specialty was categorized using a method developed by Welch et al. (primary care, surgical specialty, medical specialty, hospital based, obstetrics-gynecology, psychiatry, and unknown).[40]

Results

We identified 470,313 unique physician NPIs from the 2007-2008 carrier claims included in the SEER-Medicare data. Using Approach 1, which assessed the SEER-Medicare claims, we excluded 91,310 NPIs because there were multiple NPIs matched to one UPIN or multiple UPINs matched to one NPI. For the remaining 379,003 NPIs, we were able to match 331,763 to a UPIN (Figure 1). Using Approach 2, which identified UPINs listed in the NPI Directory, we were able to match 349,957 NPIs to a UPIN. Finally, using Approach 3, we were able to match 355,104 NPIs to a UPIN based on exact name agreement between the NPI and UPIN directories.

Approximately 48% (n=224,557) of all NPIs were linked to a UPIN within each of the three approaches (Table 1). Of the NPIs that were linked to a UPIN via all three approaches, 97.7% (n=219,315) matched to the same UPIN across all three approaches and 2.3% (n=5,179) matched to the same UPIN across two approaches only. Less than 0.1% (n=63) of the NPIs were associated with unresolvable conflicts and were not assigned a UPIN.

A further 33.9% of all NPIs (n=159,179: 45,392+49,927+63,860) were linked to a UPIN via two approaches (Table 1). Again, for the majority (n=155,098: 44,681+49,044+61,373) of these NPIs, the same UPIN was identified across the two approaches. Only 2.6% (n=4,081;

711+883+2487) of these NPIs were associated with unresolvable conflicts and were not assigned a UPIN.

For 9.5% of all NPIs (n=44,795:11,887+16,148+16,760), we were able to link to a UPIN via one approach only. Only 1.2% of these NPIs (n=545: 71+133+341) were associated with unresolvable conflicts. Finally, we found no associated UPIN via any approach for approximately 8.9% (n=41,782) of all NPIs.

An NPI-UPIN match was identified using at least one approach for 90.1% of all identified physician NPIs (n=423,842; Table 2). When classified by match confidence, 46.6% (n=219,315) of the NPIs had very strong matches (agreement across three approaches), 34.1% (n=160,277) had strong matches (agreement across two approaches) and 9.4% (n=44,250) had moderate matches (identified via only one approach). The remaining NPIs (9.9%) were ultimately classified as not having a UPIN match due to unresolvable conflicts (n=4,689; 1.0%) or no UPIN identified using any approach (n=41,782, 8.9%).

After completing the NPI-UPIN crosswalk, we examined differences in physician characteristics according to match strength (Table 3). Although more than 90% of identified physicians NPIs were linked to a UPIN, female physicians were almost twice as likely to not have had their NPI match to a UPIN, compared with male physicians (13.9% vs. 7.3%, p<0.0001, Table 3). We found a very strong or strong NPI-UPIN match for >80% of physicians in surgical, medical and hospital-based specialties, who are likely to provide first line treatment to individuals with cancer (e.g., surgical and medical oncologists). Compared to other specialists, primary care physicians were less likely to have a very strong match (42.0% vs. >46% for all other specialists) and more likely to not have a match (10.7% vs. ~7% for all other specialists, respectively, p <0.0001).

Discussion

We created a SEER-Medicare crosswalk file that links individual physician NPIs identified from the 2007-2008 SEER-Medicare claims to individual UPINs. Overall, we were able to identify a matching UPIN for more than 90% of these NPIs using information from the claims themselves, the CMS NPI and UPIN Directories and name matching techniques. This crosswalk can provide important supplemental information to the SEER-Medicare data, enhancing the ability of researchers to examine physician treatment patterns over time. This crosswalk may be of increased interest with CMS's implementation of episode-based oncology demonstration models,[20] as researchers may want to longitudinally track how individual physicians practice prior to and after the implementation of payment models.

We did identify differences in our ability to link NPIs to associated UPINs across physician characteristics. Of note, individual physicians without an NPI-UPIN match were disproportionately female. Although this may have occurred for a variety of reasons, we can speculate that women may have left their medical practices for periods of time (e.g., childrearing) or changed surnames. Additionally, in more recent years, the demographics of the physician population have undergone changes. Specifically, a larger proportion of new physicians are female and would not have been in practice prior to 2007.[41,42] Therefore,

they would not have had a UPIN. While differences in our ability to link NPIs to UPINs across physician specialty is less clear, we can speculate that the lower match rate for primary care physicians may be due to the higher proportion of women in this specialty.[43] Further, primary care physicians may trend toward part-time work or breaks in full-time employment compared with other specialties.[44,45] While we acknowledge these differences across physician characteristics, it's important to note that we were able to link a single UPIN for over 90% of the unique NPIs identified.

This crosswalk was developed specifically for use with the SEER-Medicare data. However, the crosswalk is available to all researchers, beyond those using the SEER-Medicare data and can be requested directly from the National Cancer Institute's Healthcare Delivery Research Program (http://healthcaredelivery.cancer.gov/seermedicare/considerations/crosswalk.html). SEER-Medicare users will receive the file with NPIs and UPINs encrypted in the same format as the associated Medicare files. For non-SEER-Medicare users an unencrypted version will be released. Although the NPIs identified from the SEER-Medicare claims do not include all NPIs, the process to link NPIs to UPINs that we have documented in detail in this paper can be used as a prototype for additional NPI-UPIN crosswalks in the Medicare data more broadly. Considering that the final match rate in our data was over 90%, we believe that these methods could be successfully used by others.

The National Bureau of Economic Research (NBER) has created a crosswalk file that links NPIs from the NPI Directory to possible UPINs. The accuracy of the NBER match has not been rigorously evaluated. However, there are some important differences between the NBER crosswalk and the SEER-Medicare physician NPI-UPIN crosswalk. The NBER crosswalk was created exclusively from the 2008 NPI Directory. The NBER crosswalk did not take into account the two other approaches that we used, specifically claims data were not assessed nor was there any attempt to name match. Using the NBER approach, for an NPI to be matched to a UPIN, the physician had to voluntarily report his/her UPIN on the NPI application. In addition, as noted on the NBER website, there can be multiple prior UPINs listed for each NPI in the crosswalk. The NBER Crosswalk tried to limit the match to any potential UPINs, defined as any identifier on the NPPES file that had the format of a UPIN, i.e. beginning with a letter followed by 5 digits (personal communication, Jean Roth, NBER, 2/7/13). The result is that an NPI on the NBER Crosswalk file can potentially be linked to multiple providers, unlike the 1:1 match in the SEER-Medicare physician NPI-UPIN crosswalk.

While our study creates a new resource for researchers interested in evaluating individual physician practice patterns over time, we acknowledge certain limitations. First, as with all Medicare data, physicians treating patients solely in Medicare Managed Care arrangements have not historically been required to file claims for their services to Medicare; therefore, this crosswalk will necessarily exclude these physicians. Additionally, we were not able to identify a unique NPI-UPIN match for approximately 10% of our sample. This is not unexpected and may be due to the approximately 15,000 physicians retiring each year, an additional 24,000 new physicians completing their medical training or because of missing or incomplete data on the Medicare claims and other data sources.[46]. Those physicianss who began their medical career or retired in 2007-2008 would be unlikely to have both identifiers

(i.e., retirees would only have UPINs while new graduates would only have NPIs) and, therefore, would not have an NPI-UPIN match. Finally, we included only physician identifiers for whom we could make a direct 1:1 link between NPIs and UPINs. While this conservative approach to creating the NPI to UPIN crosswalk improved our likelihood of identifying a valid match, it, by definition, removed real-life scenarios of a one-to-many match. For example, our crosswalk would necessarily exclude two physicians who billed under a group UPIN but then obtained individual NPIs. We ultimately chose an exact deterministic approach over a probabilistic approach because, as a final product, we wanted to link a single NPI with a single UPIN based on information from several data sources. While probabilistic methods may provide a more nuanced understanding of the likelihood of a true NPI-UPIN match, "exact deterministic" matching methods have been demonstrated to have high validity and reliability and have been employed successfully in multiple updates of the SEER-Medicare linked dataset.[47] Further, deterministic name matching to link an NPI to an associated UPIN was utilized as the only approach for 16,419 pairs (<3.5% of the final crosswalk). Despite this strict matching criterion, we were still able to create a unique 1:1 NPI to UPIN match for over 90% of the NPIs.

Prior to the development of this process of matching NPIs to UPINs, researchers have been limited in their ability to track individual physicians over time using claims data. The lack of a systematically developed NPI-UPIN crosswalk has presented challenges for understanding physician treatment patterns, particularly because many Medicare policy changes, including the Medicare Modernization Act,[48] that might have influenced physician practice patterns were implemented during the transition from UPINs to NPIs. The creation of this SEER-Medicare NPI-UPIN crosswalk will support continued high-quality research into individual physician treatment patterns over time.

Acknowledgments

References

1. Eberth JM, Xu Y, Smith GL, et al. Surgeon influence on use of needle biopsy in patients with breast cancer: a national medicare study. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2014; 32:2206–16. [PubMed: 24912900]

2. Krawczyk P, Ramlau R, Chorostowska-Wynimko J, et al. The efficacy of EGFR gene mutation testing in various samples from non-small cell lung cancer patients: a multicenter retrospective study. Journal of cancer research and clinical oncology. 2014

3. Pham HH, Schrag D, O'Malley AS, Wu B, Bach PB. Care patterns in Medicare and their implications for pay for performance. The New England journal of medicine. 2007; 356:1130–9. [PubMed: 17360991]

4. Reschovsky JD, Hadley J, Saiontz-Martinez CB, Boukus ER. Following the money: factors associated with the cost of treating high-cost Medicare beneficiaries. Health services research. 2011; 46:997–1021. [PubMed: 21306368]

5. Satram-Hoang S, Reyes C, Hoang KQ, Momin F, Skettino S. Treatment practice in the elderly patient with chronic lymphocytic leukemia-analysis of the combined SEER and Medicare database. Annals of hematology. 2014; 93:1335–44. [PubMed: 24638841]

6. Unroe KT, Greiner MA, Hernandez AF, et al. Resource use in the last 6 months of life among medicare beneficiaries with heart failure, 2000-2007. Archives of internal medicine. 2011; 171:196–203. [PubMed: 20937916]

7. [Accessed July 6, 2011] Medicare data available. Research Data Assisstance Center. 2011. at http://www.resdac.org/Medicare/data_available.asp

8. [Accessed Web Page, December 8, 2014] SEER-Medicare Linked Database. 2014. at http://healthservices.cancer.gov/seermedicare/

9. Who relies on Medicare? Profile of the Medicare Population. AARP Public Policy Institute; 2014. at http://www.aarp.org/content/dam/aarp/research/public_policy_institute/health/2014/who-relies-on-medicare-fact-sheet-2014-AARP-ppi-health.pdf [Accessed September 15, 2014]

10. [Accessed July 6, 2011] Health Insurance Portability and Accountability Act of 1996: General Information. 2011. at http://www.cms.gov/HIPAAGenInfo/

11. Research Data Assisstance Center: 2003. Medicare physician identifiers, UPINS, PINs and NPI numbers. at http://www.resdac.org/tools/TBs/TB-002_Medicare%20Physician%20Identifiers_508.pdf [Accessed July 5, 2011]

12. [Accessed July 6, 2011] Non-identifiable data files: Unique physician identification number (UPIN) directory. 2011. at https://www.cms.gov/NonIdentifiableDataFiles/08_UniquePhysicianIdentificationDirectory.asp

13. Borras JM, Albreht T, Audisio R, et al. Policy statement on multidisciplinary cancer care. European journal of cancer (Oxford, England: 1990). 2014; 50:475–80.

14. [Accessed April 15, 2015] PDQ Cancer Information Summaries: Adult Treatment. 2015. at http://www.cancer.gov/cancertopics/pdq/adulttreatment

15. [Accessed April 15, 2015] NCCN Guidelines for Treatment of Cancer by Site. 2015. at http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site

16. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Medical care. 2002; 40:Iv-3–18.

17. Warren JL, Harlan LC, Fahey A, et al. Utility of the SEER-Medicare data to identify chemotherapy use. Medical care. 2002; 40:Iv-55–61.

18. Virnig BA, Warren JL, Cooper GS, Klabunde CN, Schussler N, Freeman J. Studying radiation therapy using SEER-Medicare-linked data. Medical care. 2002; 40:Iv-49–54.

19. [Accessed December 8, 2014] SEER-Medicare Publications. 2014. at http://appliedresearch.cancer.gov/seermedicare/overview/publications.html

20. Backhus LM, Farjah F, Zeliadt SB, et al. Predictors of imaging surveillance for surgically treated early-stage lung cancer. The Annals of thoracic surgery. 2014; 98:1944–52. [PubMed: 25282167]

21. Brooks GA, Li L, Uno H, Hassett MJ, Landon BE, Schrag D. Acute hospital care is the chief driver of regional spending variation in Medicare patients with advanced cancer. Health affairs (Project Hope). 2014; 33:1793–800. [PubMed: 25288424]

22. DuGoff EH, Bekelman JE, Stuart EA, Armstrong K, Pollack CE. Surgical quality is more than volume: the association between changing urologists and complications for patients with localized prostate cancer. Health services research. 2014; 49:1165–83. [PubMed: 24461049]

23. Evan Pollack C, Wang H, Bekelman JE, et al. Physician social networks and variation in rates of complications after radical prostatectomy. Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2014; 17:611–8. [PubMed: 25128055]

24. Filson CP, Schroeck FR, Ye Z, Wei JT, Hollenbeck BK, Miller DC. Variation in Use of Active Surveillance among Men Undergoing Expectant Treatment for Early Stage Prostate Cancer. The Journal of urology. 2014; 192:75–81. [PubMed: 24518783]

25. Fisher KJ, Lee JH, Ferrante JM, et al. The effects of primary care on breast cancer mortality and incidence among Medicare beneficiaries. Cancer. 2013; 119:2964–72. [PubMed: 23677482]

26. Freedman RA, Vaz-Luis I, Barry WT, et al. Patterns of chemotherapy, toxicity, and short-term outcomes for older women receiving adjuvant trastuzumab-based therapy. Breast cancer research and treatment. 2014; 145:491–501. [PubMed: 24756187]

27. Lamont EB, Zaslavsky AM, Subramanian SV, Meilleur AE, He Y, Landrum MB. Elderly breast and colorectal cancer patients' clinical course: patient and contextual influences. Medical care. 2014; 52:809–17. [PubMed: 25119954]

28. Neuman HB, Weiss JM, Schrag D, et al. Patient demographic and tumor characteristics influencing oncologist follow-up frequency in older breast cancer survivors. Annals of surgical oncology. 2013; 20:4128–36. [PubMed: 23943027]

29. Onega T, Weiss J, Kerlikowske K, et al. The influence of race/ethnicity and place of service on breast reconstruction for Medicare beneficiaries with mastectomy. SpringerPlus. 2014; 3:416. [PubMed: 25140292]

30. O'Shaughnessy MJ, Jarosek SL, Virnig BA, Konety BR, Elliott SP. Factors associated with reduction in use of neoadjuvant androgen suppression therapy before radical prostatectomy. Urology. 2013; 81:745–51. [PubMed: 23465162]

31. Penn DC, Stitzenberg KB, Cobran EK, Godley PA. Provider-based research networks demonstrate greater hospice use for minority patients with lung cancer. Journal of oncology practice/American Society of Clinical Oncology. 2014; 10:e182–90.

32. Roetzheim RG, Lee JH, Ferrante JM, et al. The influence of dermatologist and primary care physician visits on melanoma outcomes among Medicare beneficiaries. Journal of the American Board of Family Medicine: JABFM. 2013; 26:637–47. [PubMed: 24204060]

33. Schneider EB, Ejaz A, Spolverato G, et al. Hospital volume and patient outcomes in hepato-pancreatico-biliary surgery: is assessing differences in mortality enough? Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract. 2014; 18:2105–15. [PubMed: 25297443]

34. Schroeck FR, Kaufman SR, Jacobs BL, et al. The impact of technology diffusion on treatment for prostate cancer. Medical care. 2013; 51:1076–84. [PubMed: 24226306]

35. Steele CB, Townsend JS, Tai E, Thomas CC. Physician visits and preventive care among Asian American and Pacific Islander long-term survivors of colorectal cancer, USA, 1996-2006. Journal of cancer survivorship: research and practice. 2014; 8:70–9. [PubMed: 24190589]

36. Veenstra CM, Epstein AJ, Liao K, Morris AM, Pollack CE, Armstrong KA. The effect of care setting in the delivery of high-value colon cancer care. Cancer. 2014; 120:3237–44. [PubMed: 24954628]

37. Wang SY, Wang R, Yu JB, et al. Understanding regional variation in Medicare expenditures for initial episodes of prostate cancer care. Medical care. 2014; 52:680–7. [PubMed: 25023913]

38. Wright AA, Hatfield LA, Earle CC, Keating NL. End-of-Life Care for Older Patients With Ovarian Cancer Is Intensive Despite High Rates of Hospice Use. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2014; 32:3534–9. [PubMed: 25287831]

39. Surveillance, Epidemiology, End, Results. (SEER) Program Research Data (1973-2007). 2010. at www.seer.cancer.gov

40. Welch WP, Cuellar AE, Stearns SC, Bindman AB. Proportion of physicians in large group practices continued to grow in 2009-11. Health affairs (Project Hope). 2013; 32:1659–66. [PubMed: 24019373]

41. [Accessed December 8, 2014] Census bureau releases equal employment opportunity tabulation that provides a profile of America's Workforce. 2012. at http://www.census.gov/newsroom/releases/archives/employment_occupations/cb12-225.html

42. [Accessed December 8, 2014] More women are doctors and lawyers than ever- but progress is stalling. The Atlantic. 2012. at http://www.theatlantic.com/sexes/archive/2012/12/more-women-are-doctors-and-lawyers-than-ever-but-progress-is-stalling/266115/

43. U.S. Department of Health and Human Services. The Physician Workforce: Projections and Research into Current Issues Affecting Supply and Demand. 2008.

44. 46% jump in number of physicians working part-time. Managed care (Langhorne, Pa). 2008; 17:19.

45. Teljeur C, Thomas S, O'Kelly FD, O'Dowd T. General practitioner workforce planning: assessment of four policy directions. BMC health services research. 2010; 10:148. [PubMed: 20525163]

46. [Accessed September 8, 2015] The Physician Workforce: Projections and Research into Current Issues Affecting Supply and Demand. 2008. at http://bhpr.hrsa.gov/healthworkforce/reports/physwfissues.pdf

47. Dusetzina, SB., Tyree, S., Meyer, AM. Linking Data for Health Services Research: A Framework and Instructional Guide. Washington, DC: Agency for Healthcare Research and Quality; 2014.

48. [Accessed September 20, 2014] Overview of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003. 2004. at http://royce.house.gov/uploadedfiles/overview%20of%20medicare.pdf
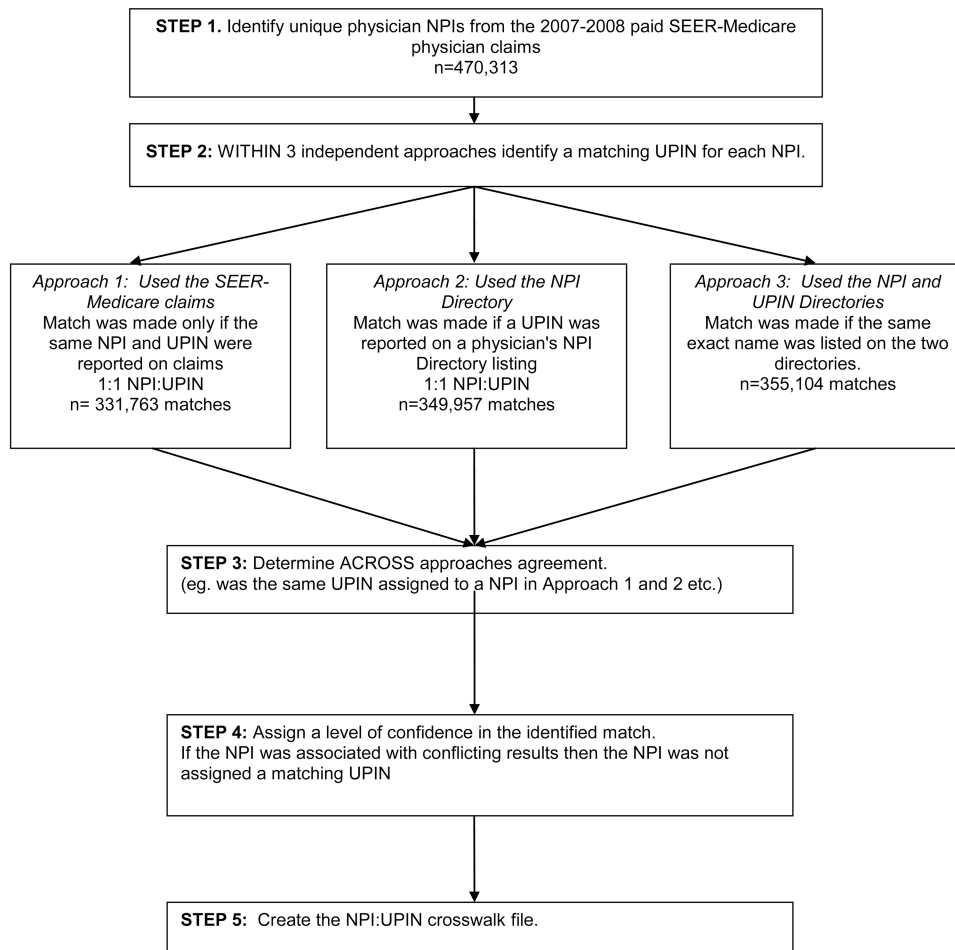
Figure 1. Process of creating a SEER-Medicare NPI-UPIN Crosswalk

NPI: National Provider Identifier; SEER: Surveillance, Epidemiology, End Results; UPIN: Unique Physician Identification Number

Table 1

Distribution of NPI-UPIN matches identified by three independent approaches and the level of across-approach agreement, for physician NPIs identified in the 2007–2008 SEER-Medicare claims data

Within-approach NPI-UPIN match					Across-approach agreement[+]			
Approach 1: SEER-Medicare claims	Approach 2: NPI Directory: listed UPIN	Approach 3: NPI and UPIN Directories: listed names	N (Total=470,313)	% of NPIs	All	Two	One source only	Unresolvable Conflicts[++]
YES	YES	YES	224,557	47.8	219,315	5,179		63
YES	YES	NO	45,392	9.7		44,681		711
YES	NO	YES	49,927	10.6		49,044		883
NO	YES	YES	63,860	13.6		61,373		2487
YES	NO	NO	11,887	2.5			11,816	71
NO	YES	NO	16,148	3.4			16,015	133
NO	NO	YES	16,760	3.6			16,419	341
NO	NO	NO	41,782	8.9				

NPI: National Provider Identifier; SEER: Surveillance, Epidemiology, End Results; UPIN: Unique Physician Identification Number

[+] Among matches that were identified within at least 2 approaches.

[++] Unresolvable conflicts defined as >1 NPI to 1 UPIN or 1 NPI to >1UPIN

Table 2

Distribution of NPI-UPIN matches by match confidence level, for physician NPIs identified in the 2007-2008 SEER-Medicare claims data.

	n	%
TOTAL	**470,313**	**100**
NPI-UPIN match confidence[*]		
Very strong	219,315	46.6
Strong	160,277	34.1
Moderate	44,250	9.4
Total with any match	*423,842*	*90.1*
No NPI-UPIN match possible		
Unresolvable conflict[+]	4,689	1.0
No UPIN found	41,782	8.9
Total NPIs not matched to UPIN	*46,471*	*9.9*

NPI: National Provider Identifier; SEER: Surveillance, Epidemiology, End Results; UPIN: Unique Physician Identification Number

[*] Very strong match=agreement across all 3 approaches; strong match=agreement across two approaches; moderate match=match made via only one approach

[+] Unresolvable conflicts defined as >1 NPI to 1 UPIN or 1 NPI to >1 UPIN

Table 3

Distribution of physician characteristics by match confidence level for physician NPIs identified in the 2007-2008 SEER-Medicare claims data

Physician Characteristics	Total		NPI-UPIN Match Confidence[1]									
			Very Strong		Strong		Moderate		Unresolvable Conflict[+]		No match		
	n	Col %	n	Row %	n	Row %	n	Row %	n	Row %	n	Row %	p-value[4]
Gender													
Male	354,462	75.4	168,101	47.4	122,915	34.7	33,970	9.6	3,755	1.1	25,721	7.3	p<.0001
Female	115,851	24.6	51,214	44.2	37,362	32.2	10,280	8.9	934	0.8	16,061	13.9	
Provider Specialty													
Primary Care	160,394	34.1	67,431	42.0	56,757	35.4	17,270	10.8	1,768	1.1	17,168	10.7	p<.0001
Surgical Specialty	71,110	15.1	34,931	49.1	23,602	33.2	6,380	9.0	672	1.0	5,525	7.8	
Medical Specialty	53,968	11.5	26,173	48.5	18,723	34.7	4,947	9.2	472	0.9	3,653	6.8	
Hospital-Based	139,558	29.7	69,048	49.5	45,652	32.7	11,509	8.3	1,348	1.0	12,001	8.6	
Obstetrics- gynecology	19,603	4.2	9,903	50.5	6,567	33.5	1,601	8.2	195	1.0	1,337	6.8	
Psychiatry	25,651	5.5	11,823	46.1	8,966	35.0	2,536	9.9	233	0.9	2,093	8.2	
N/A	29	0	6	20.7	10	34.5	7	24.1	1	3.5	5	17.2	

NPI: National Provider Identifier; UPIN: Unique Physician Identification Number

[1] Very strong match=agreement across all 3 approaches; strong match=agreement across two approaches; moderate match=match made via only one approach; conflicting results= no agreement across approaches

[2] Column percent

[3] Row percent

[4] Chi-square p-value by strength of match

[+] Unresolvable conflicts defined as >1 NPI to 1 UPIN or 1 NPI to >1UPIN