Using claims data to attribute patients with breast, lung, or colorectal cancer to prescribing oncologists

Ezra Fishman1
John Barron2
Ying Liu1
Santosh Gautam1
Justin E Bekelman3
Amol S Navathe4
Michael J Fisch5
Ann Nguyen6
Gosia Sylwestrzak1

1Translational Research, HealthCore, Inc., Wilmington, DE, USA; 2Clinical & Scientific Leadership, HealthCore, Inc., Wilmington, DE, USA; 3Radiation Oncology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA; 4Health Policy and Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA; 5Medical Oncology, AIM Specialty Health, Chicago, IL, USA; 6Oncology & Palliative Care Solutions, Anthem Inc., Woodland Hills, CA, USA

Background: Alternative payment models frequently require attribution of patients to individual physicians to assign cost and quality outcomes. Our objective was to examine the performance of three methods for attributing a patient with cancer to the likeliest physician prescriber of anticancer drugs for that patient using administrative claims data.

Methods: We used the HealthCore Integrated Research Environment to identify patients who had claims for anticancer medication along with diagnosis codes for breast, lung, or colorectal lung cancer between July 2013 and September 2017. The index date was the first date with a record for anticancer medication and cancer diagnosis code. Included patients had continuous medical coverage from 6 months before index to at least 7 days after index. Patients who received anticancer drugs during the 6 months prior to index were excluded. The three methods attributed each patient to the physician with whom the patient had the most evaluation and management (E&M) visits within a 90-day window around the index date (Method 1); the most E&M visits with no time window (Method 2); or the E&M visit nearest in time to the index date (Method 3). We assessed the performance of the methods using the percentage of the study cohort successfully attributed to a physician, and the positive predictive value (PPV) relative to available physician-reported data on patient(s) they treat.

Results: In total, 70,641 patients were available for attribution to physicians. Percentages of the study cohort attributed to a physician were: Method 1, 92.6%; Method 2, 96.9%; and Method 3, 96.9%. PPVs for each method were 84.4%, 80.6%, and 75.8%, respectively.

Conclusion: We found that a claims-based algorithm – specifically, a plurality method with a 90-day time window – correctly attributed nearly 85% of patients to a prescribing physician. Claims data can reliably identify prescribing physicians in oncology.

Keywords: alternative payment model, specialty care, plurality rule, pay for performance

Introduction

Patient–provider attribution is the process of assigning a patient to the provider – physician and/or practice – recognized as the entity most responsible for the patient’s medical care and health outcomes. Attribution originated from the methods used by the Centers for Medicare and Medicaid Services (CMS) and private payers to assign patients to primary care providers.1–4 Attribution has grown in importance in the current health care environment, where both private payers and CMS have increased the use of alternative payment models that put providers at risk for the costs and outcomes of their patients.2,5 These reimbursement strategies rely on payers to identify the provider(s) most responsible for the care of any given patient.4 Inaccuracies in this identification...
ascribe providers with responsibility for patients not actually under their management, which can result in a misalignment of incentives and a lack of fairness when used in value-based payment models.6,7

There is currently a dearth of knowledge about attribution in specialty care, especially for individual physicians. The few available studies on provider attribution in cancer care have focused on the assignment of patients to practices rather than to individual physicians.8,9 Attribution to individual physicians is important because alternative payment models seek to evaluate physicians on the basis of value.10,11 Moreover, physician-level differences in treatment patterns explain a large portion of variation in utilization and spending, even when controlling for the practice or hospital where the treatment takes place.10–13 No study, to our knowledge, has described and assessed patient attribution to individual physicians in oncology. The ability to identify a prescribing physician from administrative claims data is especially important because claims data are the most available way to track practice patterns of physicians and link those patterns to their cost implications.

The simplest way to attribute patients to providers would be to assign the administering provider on the index data as the attributed provider; however, this approach is not likely to be accurate for several reasons. First, the index claim for anticancer drugs is sometimes a pharmacy claim, where the prescribing provider is not necessarily listed. Second, even in medical claims, the index administering provider identifier often maps to an institution and not a person, as chemotherapy can be administered in a facility that is separate and independent from the prescriber’s practice. A final technical challenge relates to the ability to distinguish among individual physicians within a practice, a difficulty that arises when oncologists cross-cover for one another and sign-off on chemotherapy orders that were chosen by their colleagues. Thus, the correct assignment of cancer patients to practices (or health care delivery systems) may not necessarily lead to accurate attribution of patients to individual physician prescribers; therefore, programs and policies aimed at influencing individual physician behavior need to employ attribution methods that are effective at the individual physician level.

The aim of this study was to characterize different claims-based methods that identify the prescribing physician for each patient receiving anticancer drugs. As anticancer drugs are a major cost driver in oncology, with significant variation in value across regimens, they are an important locus for improvement in the quality and value of cancer care.14,15 These claims-based methods attribute patients to individual physicians, rather than practices or health care systems. The performance of each method was measured as the attributed percentage, which represented the proportion of patients in the study cohort who were assigned a prescribing physician, and by the positive predictive value (PPV) of the attribution relative to the subgroup of patients with available physician-reported data on patient(s) they treat.

**Methods**

**Data source**

We used administrative claims data from the HealthCore Integrated Research Environment (HIRE®) for information on diagnoses, utilization of cancer treatment, and rendering provider identifiers at the claim line level. The HIRE is a repository of medical and pharmacy claims data for ~40 million members managed by 14 commercial health plans geographically dispersed across the United States.

To validate the claims-based algorithms, claims and eligibility data were linked at the patient level to information reported by physicians through an online portal built especially for an oncology program that enhances reimbursement to prescribing oncologists for care coordination.15,16 The physician-reported information included patient name (masked before the data were made available to researchers), date of birth, date of treatment, the prescribed anti-cancer drug regimen, the prescribing physician’s National Provider Identifier (NPI), and a number of clinical details unavailable in claims. Physicians identified their patient(s) as part of inputting data into the portal, thus establishing “gold standard” physician–patient dyads.6 Physicians needed to report these data in order to identify whether the prescribed regimens would qualify the physicians for enhanced care coordination reimbursement from the 14 participating health plans.

This study was conducted in full compliance with relevant provisions of the Health Insurance Portability and Accountability Act. As researchers only used the analytical file derived from a limited data set to perform the analyses as defined by the Privacy Rule 45CFR 164.514(e), no waiver of informed consent or exemption was needed from an institutional review board.

**Study population, inclusion criteria, and exclusion criteria**

We aimed to use claims to identify a set of patients that would approximate the patient population subject to a value-based reimbursement program in oncology. We identified patients who had a claim for an anticancer drug and a claim with diagnosis code for breast, lung, or colorectal lung cancer on
the same date of service between July 2013 and September 2017. The first service date with an anticancer medication and a diagnosis for one of the three cancer types was defined as the index date. Anticancer drugs were identified using Current Procedural Terminology (CPT)/Healthcare Common Procedure Coding System (HCPCS) codes included in medical claims and generic product identifier codes included in pharmacy claims data (Supplementary materials, Tables SA1–SA4).

We applied a hierarchical method for patient selection, which facilitated the indexing of patients according to the anticancer drugs they received and linking them with the most identifiable prescribers. First, we identified and selected patients receiving injectable agents. From the remaining patients, we identified and selected those on oral agents (excluding hormonal therapy), and finally, we identified additional patients with breast cancer who received only oral hormonal therapy. We prioritized injected agents because they are generally administered at physicians’ offices, hospitals, or special centers and billed under medical, rather than pharmacy, benefits. Many payer initiatives and research studies will be interested in capturing complete anticancer treatment regimens, which often include multiple administrations over weeks or months, rather than a single administration on a single day. To ensure we captured complete regimens, patients were required to have continuous enrollment in medical and pharmacy benefits from 6 months before the index date to 7 days after the index date. We excluded patients who had claims for anticancer therapy in the 6 months prior to index date because we likely did not capture the beginning of their course of anticancer therapy. We additionally excluded patients with multiple cancer types (eg, both breast and lung cancer), because the complexity of their disease profile would make it difficult to discern and associate treatment regimens and cancer types.

**Attribution of prescribing physicians and patients**

We considered three methods for using claims to assign each patient to an attributed physician. The first, shown in the top panel of Figure 1, is the plurality method with a 90-day time window; the second removes the time window but otherwise

![Figure 1](https://example.com/figure1.png)  
*Figure 1* Example of methods of attributing patients to physicians.
is the same as the first method; the third, shown in the bottom panel of Figure 1, is the nearest visit method. In all cases, providers are identified using the NPI appearing on claims. Each NPI uniquely identifies a provider – a physician or an institution – over time.

In the plurality method (both with and without a time window), the physician appearing on the largest number of claims during a particular time window around the index date is assigned as the attributed physician. In this study, we only considered claims from physicians with codes indicating a specialty in oncology. Oncologists were identified on the basis of CMS specialty codes associated with a given NPI: 82 (hematology), 83 (hematology/oncology), 90 (medical oncology), 91 (surgical oncology), 92 (radiation oncology), 94 (interventional radiology), or 98 (gynecological oncology). We only considered claims that had only CPT/HCPCS codes indicating office visits for evaluation and management (E&M) on the basis of CPT/HCPCS code 99201–99499 on the claim line.

In the first method, the plurality method with a 90-day time window claims contributed to the calculation of a plurality if they had service dates from 30 days prior to index date (date of anticancer therapy initiation) to 60 days after the index date. The second method removed the time restriction. The absence of a time restriction in the second method generated a large number of situations where more than one physician appeared on the same number of claims for a given patient. These “ties” were broken by assigning the physician with a service date closest to the patient’s index date as the attributed physician.

In the third method – the nearest visit method – the oncologist appearing on a claim for an E&M office visit nearest in calendar time to the index date is assigned as the attributed physician. The “nearest in calendar time” could be before or after the index date. This method assigns the rendering provider to a patient based on the claim with the smallest absolute difference between the index date and the date of E&M services. In this attribution method, oncologists were identified with the same specialty codes as in the plurality method.

Assessing validity of attribution methods
We used two separate measures to validate the attribution methods. The first measure was the proportion of patients in the study cohort who had a prescribing (attributed) physician. This measure is important because researchers and health plans want to be able to maximize the number of patients used to describe and assess physician behavior. The second measure was the PPV, which represents the percentage of patients who were attributed the “correct” prescribing physician, as verified with physician-reported data. A high attributed percentage and a high PPV from our claims-based attribution method would signal strong grounds to apply the claims-based method to patients and physicians who might not show up in the physician-reported data at all.

Because not all physicians reported data through the online portal or participated in the enhanced reimbursement program, only a subset of attributed patients had the opportunity to have their prescribing physician verified. We selected the subgroup of patients who appear in the physician-reported data within ±30 days from their claims-based index date, to ensure that a patient’s appearance in the physician-reported data was for the index cancer and the index treatment regimen. For each patient, we retrieved the NPI of the prescribing physician who submitted the data into the portal. (Of the patients appearing in the physician-reported data in this time window, 97% were associated with exactly one reporting physician; the remaining 3% were randomly assigned a single physician from among those reporting on that patient.) Then we compared the physician–patient dyads identified from claims data to the information from the physician-reported data. When the claims-based dyad agreed with the patient-ID/prescribing-provider-ID shown in the physician-reported data, the attribution was considered validated.

To summarize, the denominator for the PPV was the number of patients who 1) received an attributed NPI using the claims-based method under investigation and 2) were represented in the physician-reported data in the period beginning 30 days prior to the index date and ending 30 days after the index date, suggesting the possibility that their attributed NPI could be validated. The numerator for the PPV was the number of patients 1) who contributed to the denominator and 2) whose attributed NPI was the same as the prescriber’s NPI in the physician-reported data.

Results
Study population
Our queries of the HIRE yielded a total of 141,211 patients with breast, lung, or colorectal cancer who were receiving anticancer treatment during the period of interest. Upon excluding patients who did not meet the health plan enrollment criteria and those who received cancer drugs in the 6 months before the index date, a total of 70,641 patients remained available for assignment to attributed physicians, as shown in Figure 2. These 70,641 patients constituted the study population for all attribution methods and the denominator for the attributed percentage in the plurality method and
nearest visit methods. Patients excluded from the study did not differ significantly from the 70,641 included patients in the distribution of age, sex, or cancer type (Supplementary materials, Table SB1).

**Percent of patients attributed to physician**

From the study population, 65,379 and 68,440 patients were attributed a prescribing physician for the 90-day window and no time window, respectively, of the plurality method. The resulting attributed percentages were 92.6% and 96.9% (Figure 2). The nearest-visit method also attributed a prescribing physician to 68,440 (or 96.9%) of patients. The percentage of patients attributed a physician were the same for nearest-visit method as for the second application of the plurality method because neither placed a time restriction on when the oncologist E&M visits could occur in relation to the start of anticancer treatment.
PPV for plurality and nearest date to index methods

Of the attributed patients, only a minority appeared in the physician-reported data: 18,312 for the 90-day time window plurality method, 18,887 for the no time window plurality method, and the same 18,887 for the nearest-visit method. These constitute the denominators of the PPVs. Patients included in the PPV calculation had seen a similar number of oncologists, on average, as those excluded from the PPV calculation (Supplementary materials, Tables SB2 and SB3).

The respective PPVs for the three approaches were 84.4% (plurality method with 90-day window), 80.6% (plurality method with no time restriction), and 75.8% (nearest-visit method), as shown in Figure 2. The results did not differ greatly by cancer type (breast, colorectal, and lung) and are shown in Supplementary materials, Tables SC1–SC3.

Discussion

In this study, we examined the relative performance of three methods for attributing a patient to an individual oncologist. Attributing the physician with whom the patient had the most E&M visits in a 90-day period around the start of anticancer treatment resulted in the lowest percentage of the cohort getting an attributed prescribing physician (92.6%), but the highest PPV (84.4%). Eliminating the 90-day time window raised the attributed percentage to 96.9%, but reduced the PPV to 80.6%. Assigning the physician with the E&M visit nearest in time to the start of anticancer treatment further reduced the PPV to 75.8%. Based on the highest PPVs in this study, the plurality method with a time-limited window (relative to the index date) appears to be the most accurate way to identify E&M visits and to identify physicians ordering anti-cancer treatments without significantly affecting the ability to attribute a large proportion of patients. Eliminating the time window appeared to increase slightly the number of patients attributed to a physician, but at the cost of losing accuracy. The nearest-visit method performed worst in terms of accuracy, with no gain in the proportion of patients attributed a physician.

This work should be of value to researchers seeking to identify an index physician in specialty care settings. It should also be helpful to payers that want to determine the viability of payment strategies that depend on assigning patients to a particular physician responsible for their care and outcomes and help to tie the physician’s compensation to those outcomes. For attribution to specialists responsible for management of a particular patient, some version of a plurality rule with emphasis on E&M claims, limited to providers with particular specialty codes, is likely to fare better than a method that attempts to find a single claim that provides all the necessary information. It is likely that programs and research that target decision making about physicians would need to use algorithms that are as accurate as possible without excluding too many patients, suggesting a preference for the plurality method with time-limited window.

There is little existing literature assessing the accuracy of individual physician attribution, and a recent review highlighted the importance of clinician attestation in validating attribution methods. This study uses this strategy as a gold standard for validation, specifically employing physician-reported identification of their patients. A report on attribution of cancer treatment episodes to practices (rather than individual physicians) compared two attribution approaches and found 83% concordance, but it lacked a standard against which to evaluate the approaches. Our finding of 85.6% PPV using the plurality method with a 90-day time window compares favorably. An integrated health care organization reported that several of its primary care physician attribution methods were able to assign ~90% of their members to primary care physicians, similar to our 93% attributed percentage in the plurality method with 90-day time window.

Although value-based and other reimbursement models that rely on attribution are most common in primary care, they are increasingly important in specialty care as well. For example, value-based programs focusing on oncology might seek to identify the physician most responsible for managing a given patient’s chemotherapy treatment. Such a program could be run separately from any value-based program in primary care. The Oncology Care Model instituted by the CMS pays participating practices a per-patient per-month fee for each patient who starts a course of chemotherapy. Practices qualify for the payments by providing a variety of patient-centered services, some of which are challenging to bill on a fee-for-service basis.

One of the strengths of this study is that we have “gold standard” data on patients whom the prescribing physicians themselves reported as their patients, against which to test our claims-based attribution method. This gold standard covers only physicians who enter information into a portal set up for a specialized oncology reimbursement program, not all physicians reimbursed by the participating 14 health plans. More generally, other payers, including CMS, do not have access to this type of physician-reported data. Such a gold standard was not available to a prior study of specialty attribution. As a consequence, a generalizable value-based payment system will have to rely primarily on claims data,
underscoring the value of the claims-based attribution algorithms we have presented.2,4

Limitations
Cohort exclusion criteria reduced the size of the cohort from 141,211 to 70,641. Although exclusions of this magnitude are not unusual in studies of chemotherapy regimens,21 they potentially limit the study’s generalizability. The largest excluded group was that of patients who had a claim for chemotherapy before the index date. These patients were mostly those identified near the beginning of the intake period. We found no significant differences between the characteristics of excluded patients and those of included patients.

Another limitation is that the patients who appear in the physician-reported data might differ from the larger target cancer population. We found that the average number of separate oncologists seen by the patients found in the physician-reported database was similar to the average number of oncologists seen by attributed patients not found in the physician-reported database, suggesting that the accuracy of the attribution methods (ie, the PPV) is likely to be similar across the two groups.

Another limitation is that our results must be viewed against some of the inherent deficiencies of claims data including miscoding and associated inaccuracies. However, these deficiencies would be faced by any payer trying to implement a value-based payment model at scale, and by any researcher working with claims data.

Conclusion
With the growth in programs rewarding participating physicians with incentives for prescribing evidence-based and guideline-driven treatment regimens, there will be a greater need for the accurate attribution of patients to the physicians directing a majority of their care. This study found a claims-based algorithm – the “plurality method” – that attributed over 90% of patients to a prescribing physician, doing so with about 85% accuracy, suggesting that claims data can be used reliably to identify prescribing oncologists.

Acknowledgments
This study was funded by Anthem Inc., which had no role in the conduct of the study or the decision to publish the results. The authors acknowledge Bernard Tulsi for writing and editing support, and Ramya Avula for programming support.

Disclosure
Ezra Fishman, John Barron, Ying Liu, and Gosia Sylwestrzak are employees of HealthCore, Inc., a wholly owned, independently operated subsidiary of Anthem, Inc. Santosh Gautam at the time of this study was employed by HealthCore, Inc. Michael J Fisch is an employee of AIM Specialty Health, a wholly owned subsidiary of Anthem, Inc. Ann Nguyen is an employee of Anthem, Inc. Amol S Navathe reported grants from Hawaii Medical Service Association, Anthem Public Policy Institute, Cigna, and Oscar Health; personal fees from Navvis and Company, Navigant Inc., Lynx Medical, Indegene Inc., Sutherland Global Services, and Agathos, Inc.; personal fees and equity from NavaHealth; speaking fees from the Cleveland Clinic; serving as a board member of Integrated Services Inc. without compensation, and an honorarium from Elsevier Press, none of which are related to this manuscript. The authors report no other conflicts of interest in this work.

References
1. Pope, G. Chapter 7: Attributing Patients to Physicians for Pay for Performance. In: Crompton J, Trisolini MG, Pope G, et al editors. Pay for Performance in Health Care: Methods and Approaches. Park, NC: RTI Press; 2011:7:181–201. Available from: Accessed at https://www.rti.org/pubs/bk-0002-1103-mitchell.pdf, Accessed January 24, 2019.
2. Higgins A, Zeddies T, Pearson SD. Measuring the performance of individual physicians by collecting data from multiple health plans: the results of a two-state test. Health Aff. 2011;30(4):673–681.
3. Lasko TA, Atlas SJ, Barry MJ, Chueh HC. Automated identification of a physician’s primary patients. J Am Med Inform Assoc. 2006;13(1):74–79.
4. Pham HH, Schrag D, O’Malley AS, Wu B, Bach PB. Care patterns in Medicare and their implications for pay for performance. N Engl J Med. 2007;356(11):1130–1139.
5. Robinson JC. Value-based physician payment in oncology: public and private insurer initiatives. Milbank Q. 2017;95(1):184–203.
6. National Quality Forum April 2018 report, page 28. Available from: https://www.qualityforum.org/Projects/a-b/Attribution/Draft_Report_for_Comment.aspx. Accessed January 24, 2019.
7. Atlas SJ, Chang Y, Lasko TA, Chueh HC, Grant RW, Barry MJ. Is this “my” patient? Development and validation of a predictive model to link patients to primary care providers. J Gen Intern Med. 2006;21(9):973–978.
8. Abt Associates. First Annual Report from the Evaluation of the Oncology Care Model: Baseline Period. February 1, 2018; Contract #HHSM-500-2014-00026I T0003 for the Centers for Medicare and Medicaid Services (CMS). 2018.
9. Huckfeldt PJ, Chan C, Hirshman S, et al. Attribution of Episodes to Practices. In Specialty Payment Model Opportunities and Assessment; 2015. RAND Corporation. Available from: http://www.jstor.org/stable/10.7249/j.ctt14bs4v3.13. Accessed January 24, 2019.
10. Mehrrota A, Adams JL, Thomas JW, McGlynn EA. The effect of different attribution rules on individual physician cost profiles. Ann Intern Med. 2010;152(10):649–654.
11. Tsugawa Y, Jha AK, Newhouse JP, Zaslavsky AM, Jena BA. Variation in physician spending and association with patient outcomes. JAMA Intern Med. 2017;177(5):675–682.
12. Jagsi R, Griffith KA, Heimburger D, et al. Choosing wisely? Patterns and correlates of the use of hypofractionated whole-breast radiation therapy in the state of Michigan. Int J Radiat Oncol Biol Phys. 2014;90(5):1010–1016.
13. Lipitz-Snyderman A, Sima CS, Atoria CL, et al. Physician-Driven variation in Nonrecommended services among older adults diagnosed with cancer. JAMA Intern Med. 2016;176(10):1541–1548.
14. Bach PB. Costs of cancer care: a view from the centers for Medicare and Medicaid services. *J Clin Oncol*. 2007;25(2):187–190.
15. Gautam S, Sylwestrzak G, Barron J, et al. Results from a health insurer’s clinical pathway program in breast cancer. *J Oncol Pract*. 2018;14:e711–e721.
16. Malin J, Nguyen A, Ban SE, et al. Impact of enhanced reimbursement on provider participation a cancer care quality program and adherence to cancer treatment pathways in a commercial health plan. *J Clin Oncol*. 2015;33(15 Suppl):6571–6571.
17. Agiro A, Ma Q, Acheson AK, et al. Risk of neutropenia-related hospitalization in patients who received colony-stimulating factors with chemotherapy for breast cancer. *J Clin Oncol*. 2016;34(32):3872–3879.
18. Huckfeldt PJ, Chan C, Hirshman S, et al. Specialty payment model opportunities and assessment: oncology model design report. *Rand Health Q*. 2015;5(1):11.
19. HealthPartners technical report. Available from: https://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/cntrb_031064.pdf. Accessed January 24, 2019.
20. Chang E, Buist DS, Handley M, Pardee R, Gundersen G, Reid RJ. Physician service attribution methods for examining provision of low-value care. *EGEMS*. 2016;4(1):1276.
21. Winn AN, Keating NL, Troedson JG, Basch EM, Dusetzina SB. Spending by commercial insurers on chemotherapy based on site of care, 2004–2014. *JAMA Oncol*. 2018;4(4):580–581.