Published in final edited form as:

Med Care. 2018 March ; 56(3): e16–e20. doi:10.1097/MLR.0000000000000709.

Can Claims Data Algorithms Identify the Physician of Record?

Eva H. DuGoff, PhD, MPP¹, Emily Walden, MS², Katie Ronk, BS³, Mari Palta, PhD¹, and Maureen Smith, MD, PhD^{1,3}

¹Department of Population Health Sciences, University of Wisconsin-Madison, Madison, Wisconsin

²Department of Economics, University of Wisconsin-Madison, Madison, Wisconsin

³Health Innovation Program, University of Wisconsin-Madison, Madison, Wisconsin

Abstract

BACKGROUND—Claims-based algorithms based on administrative claims data are frequently used to identify an individual's primary care physician (PCP). The validity of these algorithms in the US Medicare population has not been assessed.

OBJECTIVE—To determine the agreement of the PCP identified by claims algorithms with the PCP of record in electronic health record data.

DATA—Electronic health record and Medicare claims data from older adults with diabetes.

SUBJECTS—Medicare fee-for-service beneficiaries with diabetes (N=3,658) ages 65 and older as of January 1, 2008 and medically housed at a large academic health system.

MEASURES—Assignment algorithms based on the plurality and majority of visits and tie breakers determined by either last visit, cost, or time from first to last visit.

RESULTS—The study sample included 15,624 patient-years from 3,658 older adults with diabetes. Agreement was higher for algorithms based on primary care visits (range: 78.0% for majority match without a tie breaker to 85.9% for majority match with the longest time from first to last visit) than for claims to all visits (range: 25.4% for majority match without a tie breaker to 63.3% for majority match with the amount billed tie breaker). Percent agreement was lower for non-White individuals, those enrolled in Medicaid, experiencing a PCP change, and more than 11 physician visits.

CONCLUSIONS—Researchers may be more likely to identify a patient's PCP when focusing on primary care visits only, however, these algorithms perform less well among vulnerable populations and those experiences fragmented care.

Corresponding Author: Eva H. DuGoff, PhD, MPP, 610 Walnut St, 701 WARF, Madison, WI 53726, Tel. 608-263-2082, eva.dugoff@wisc.edu.

Funding Disclosure: This study was supported by R21 HS021899 and R01 HS018368 under the Agency for Healthcare Research and Quality (Smith) and the National Institute of Mental Health under Ruth L. Kirschstein National Research Service Award T32 MH18029 (Walden). Additional support was provided by the Health Innovation Program, the UW School of Medicine and Public Health from The Wisconsin Partnership Program, and the Community-Academic Partnerships core of the University of Wisconsin Institute for Clinical and Translational Research (UW ICTR) through the National Center for Advancing Translational Sciences (NCATS), grant UL1TR000427.

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

Keywords

Continuity of care; care coordination; multimorbidity

INTRODUCTION

Continuity of care, a coherent and connected experience of care, is considered a corner stone of primary care practice and particularly important for adults with multiple chronic conditions who are most likely to experience fragmented care.¹⁻³ Administrative claims data sets are commonly used to describe patterns of continuity and evaluate care coordination interventions.⁴⁻⁷ While claims data offer researchers many advantages, such as detailed information on large samples and affordability, they typically do not identify the patient's primary care physician (PCP). Researchers have sought to overcome this limitation by developing various algorithms to impute this information.⁸⁻¹⁰ However, the validity of these algorithms in the US Medicare population and among vulnerable subpopulations has not been rigorously assessed.

Researchers have developed a number of algorithms for assigning patients to physicians and physician groups. In the continuity of care literature, the Usual Provider of Care Index is often implemented by defining the physician who delivered the plurality of physician visits as the primary physician.^{4,11-14} An important limitation of the plurality algorithm is that it does not work well if there is a tie between two physicians. Another approach is to define the primary physician as the physician who accounted for the majority of physician visits. However, this approach does not work well if there is a tie or if a single physician does not account for more than half of the visits. Other approaches have suggested by researchers interested in attributing patients to physicians have included focusing on the largest amount billed and using tie breaker rules to maximize attribution.^{8,15,16}

Despite the importance of identifying accountable physicians, the accuracy of claims-based physician assignment algorithms has received limited attention. A 2007 study using adult diabetics in Canada found that the best algorithm relying on the plurality of visits correctly identified a patient's self-reported PCP in 82.6% of cases. The external generalizability of this study to the US Medicare population is limited due to the survey's low response rate (<40%), small sample of adults living in a single geographic region (Ontario), on average younger study sample than the US Medicare population, and differences in the Canadian and US health care systems.

Other studies have documented the implications of different assignment methodologies for purposes of determining quality measures or pay for performance bonuses. Thorpe and colleagues compared two algorithms in the context of a Midwestern academic medical center. They found that over 30% more patients were assigned to physician groups by using a plurality algorithm than by using a proprietary statewide public reporting collaboration algorithm.¹⁶ The authors also found differences in the patients assigned to providers with respect to sex, chronic diseases, and predicted utilization. Using commercial health plan data, Mehrotra and colleagues compared 12 different assignment algorithms and found that

depending on the algorithm between 20 to 69% of care episodes could be assigned to a physician.⁹

The primary objective of this study was to evaluate the concordance of primary care physician attribution algorithms applied to administrative claims data to the designated primary care physician listed in the patient's electronic health record (EHR). We study this issue using a comprehensive data set of older adults with diabetes medically housed at a Midwestern health care system.. In secondary analyses, we examined whether concordance varied by health care utilization and among vulnerable populations. We studied this issue in the context of older adults with diabetes and other chronic conditions. We focused on individuals with diabetes because this condition is common, costly, and often accompanied by comorbidities.¹⁷

METHODS

Study Population and Data

We used electronic health record data from a large academic health system located in the Midwest to identify Medicare fee-for-service beneficiaries ages 65 and older as of January 1, 2008 who had a prior diagnosis of diabetes. We linked the academic health system data to Medicare fee for service claims from 2008 to 2013 to construct a patient-year dataset. These data contain beneficiary demographic characteristics, date of death, chronic condition diagnosis history, diagnosis codes, procedure codes, dates of service, type of visit, and physician identifiers.

Our study sample included 4,136 patients (19,760 patient-years) who were at least 65 in 2008, enrolled in Medicare Parts A and B, and medically housed within the health system in 2008 or 2009. We defined patients as medically housed within the health system as those who received primary care at the health system and had a designated primary care physician within the health system in 2008 or 2009. We exclude patient-years where an individual had 3 or fewer evaluation and management visits in any year because claims-based algorithms are highly sensitive to cases with few visits. Our final sample included 3,658 patients (15,624 patient-years).

The University of Wisconsin-Madison Institutional Review Boards determined that this study was not human subjects research.

Physician Specialty and Primary Care Physician of Record

We identified the patient's designated PCP using the health system's EHR. The EHR maintains a record of an individual's PCP of record over time. Within the health system, physicians in family medicine, internal medicine, pediatrics, and geriatrics may serve as the PCP of record. In rare cases, physicians within the health system with specialties such as infectious disease have undergone the system approval process to be designated a primary care physician.

From the health system's perspective, the PCP of record is a patient's primary physician who is held accountable within the health system for their patient panel's performance on

quality measures. In addition, the PCP of record field is corroborated by the patient at each encounter with the health system, and can be updated by the patient. Per institutional protocols, patients are asked at check-in: “Is your primary care physician still Dr. X?”. When a physician leaves the health care system, patients are assigned a new primary care physician or more identify a different physician. In cases where a patient had multiple designated physicians in a given year, we selected the physician who was designated for the most days in that calendar year.

It is important to note, patients may designate their PCP of record to be a physician not employed by the health system. In 2011, the mid-point of the study period, 77% of designated PCPs were within the health system.

We defined a physician as a PCP if the majority of the physician’s evaluation and management visits were attributed to one of the following specialty codes: family practice, general practice, geriatric medicine, or internal medicine or if that physician was listed in the electronic health record (EHR) as a primary care physician of record.

Other Measures

We used previously published claims-based algorithms, majority and plurality rules, which determine an individual’s primary physician using her calendar year physician visits.^{8,10} The plurality rule designates the physician with whom the patient had the most visits in a year, while the majority physician designates the physician with whom the patient had greater than 50% of visits. As specified, the majority and plurality algorithm may leave a patient without an assigned PCP, and the plurality algorithm may identify multiple physicians. In these cases, the patient-year was retained in the denominator and classified as a non-match. To reduce the number of patients with unassigned physicians or multiple physicians, we also tested three alternative tie breaker rules that assigned the patient to the physician who performed the last visit, billed the greatest share of health care expenditures, or had the longest duration of time from first to last visit for that calendar year.

Older adults with diabetes may see multiple primary care and specialist physicians for their routine care needs.³ We examined claims-based continuity algorithms using a broad and narrow definition of eligible physician visits. In our broad definition, we focused on evaluation and management claims occurring with any primary care and specialist physician (Any Visit). In our narrow definition, we define focused on evaluation and management claims occurring with only PCPs (Primary Care Visits).

We used the beneficiary summary file to identify beneficiary race (White or non-White), age, and Medicaid enrollment (any versus none). We classified individuals by number of chronic conditions (less than 2 versus 3 or more) in 2008 using the Elixhauser algorithm.¹⁸ We consider the counts of chronic conditions (including diabetes).

Analysis

All analyses were conducted in SAS version 9.4. We examined the patient sample as of January 1, 2008, and physician panel size for the 2008 calendar year. The unit of analysis was the patient-year. Concordance is measured as the percent of patient-years where the

designated PCP in the EHR and claims-based algorithms identified the same physician. The denominator included all eligible patient-years (patient years with 4 or more physician visits or primary care visits, as appropriate) regardless of whether the claims-based algorithm identified a physician.

Based on the Aday-Andersen health behavior model, we expected that patient visit patterns may vary by individual-level enabling (Medicaid status), predisposing characteristics (race), and health characteristics (number of chronic conditions). We also examined how concordance varied by health care utilization including proportion of visits within the health system and PCP change.^{19,20}

In bivariate analyses, we use the McNemar's test for paired data to assess whether the observed agreement between the majority and plurality algorithms significantly differed. We define a significant difference as p-values less than or equal to 0.05.

In our main analysis, patient subgroups who are more likely to meet the minimum number of required visits (>3 E&M visits) could unduly influence the results. To assess the sensitivity of our results, we also conducted the analysis where each person-year was inversely weighted by the individual's total person-year observations. We also the sensitivity of our results to our unit of analysis: patient-year. We also tested a patient-level analysis using a cross-sectional approach aggregating data from 2012 and 2013.

RESULTS

The study sample included 3,658 older adults with diabetes (15,624 patient years) receiving primary care in a large academic health system (Table 1). Over half of patients in the sample were female (54%) and 96% were White. As of 2008, less than 10% of the sample received Medicaid, the average number of chronic conditions was 3.6, and the average patient age was 75.9 years. Over the five-year study period, 30.4% of the sample died and 31.7% changed their PCP of record.

We observed evaluation and management visits with 5,761 unique physicians between 2008 and 2013; 1,316 of these physicians were classified as PCPs. The vast majority of physician visits occurred within the health system: 74.9 percent of all E&M visits and 87.2 percent of PCP visits took place within the health system from 2008 to 2013. A subset of PCPs (N = 505) were identified as at least one patient's designated physician for the plurality of days in at least one year during the study period. In 2008, the average PCP panel size among the study sample was 10.6.

Overall, administrative continuity algorithms limited to PCP visits had higher concordance rates than when all physicians visits were included (Table 2). When limited to PCP visits, agreement ranged from 78% (majority match without a tie breaker) to 86% (majority match with the longest time span tie breaker). When using the all physician visits, percent agreement ranged from 25% (majority match without a tie breaker) to 63% (majority match with the amount billed tie breaker).

Table 3 presents the percent agreement between the plurality and majority match algorithms without a tie breaker and the designated PCP of record by patient characteristics. Across all patient subgroups, we found that the plurality algorithms had significantly higher concordance rates than the majority algorithm, and that the analysis limited to primary care physician visits generally had higher concordance than when all visits were included. In analyses limited to primary care visits, the plurality algorithms did not identify over 20% of designated PCPs of patient-years classified as Non-White, Medicaid, patients with more than 10 physician visits, and those experience a PCP change during the year had lower concordance. The majority algorithm did not identify 30% or more of designated PCPs for these patient subgroups.

Among patient-years where 100% of their primary care visits were within the health system, the claims-based algorithms correctly identified a patient's PCP of record in 88% of cases using the plurality algorithm and 82% of cases using the majority algorithm. Among patient-years where 100% of all visits were within the health system, the claims-based algorithms only identified a patient's PCP of record in 56% of cases using the plurality algorithm and 25% of cases using the majority algorithm.

To examine whether health care utilization patterns may be driving differences in concordance by race and Medicaid status, we examined the number of physician visits and proportion of within system visits stratified by race (Appendix Table 1) and Medicaid status (Appendix Table 2). We find that non-White patients have a somewhat lower rate of within system PCP visits, but a higher rate of within system visits overall; White and non-White patients have similar numbers of physician visits and primary care visits. We found similar results by Medicaid status.

Analyses of the percent agreement by patient characteristics using alternative tie breakers and no tie breakers did not differ qualitatively from the main analysis. Sensitivity tests inversely weighting the person-year observations and using a cross-sectional patient-level approach (Appendix Table 3) also returned results similar to the main analysis.

DISCUSSION

We find claims-based assignment algorithms misclassify a patient's PCP of record from 14% to as much as 75% of the time, depending on the algorithm. Misclassification rates were higher for vulnerable subgroups: non-Whites and individuals dually enrolled in Medicare and Medicaid. Misclassification rates were also higher among individuals mostly likely to experience fragmented care: those using multiple health system, experiencing 11 or more physician visits, and those who changed PCPs during the year. Unexpectedly, majority and plurality match algorithms performed similarly when stratified by number of chronic conditions.

Researchers should use claims-based assignment algorithms to identify primary care physicians with caution. Consistent with a previous Canadian study, we found that these algorithms may misclassify a substantial proportion of individuals. This study provides new evidence that these algorithms perform less well in vulnerable populations.⁸ These findings

extend previous work by examining variations in concordance by patient characteristics. We find claims-based algorithms misidentify the PCP of record more often among non-White and Medicaid subgroups. While lower concordance in these groups could be interpreted as an indication that claims-based algorithms do not work as well in these subgroups. We think a more likely interpretation is that lower concordance is an indicator that these subgroups either access the health care or are referred through the health system by their providers in a systematically different way than other patient populations.

We were surprised to find that over 30% of the sample changed PCPs during the study period. A study in the Veterans Administration have reported that 9% of patients experienced a PCP change due to provider exits from the health care system.²¹ The higher rate of PCP change may be a function of the longer study period as well as the inclusion of residency clinics in our sample. We do not know based on these data if the patient or health system initiated the change in designated PCP. However, we know that patients at least confirmed their PCP's name at any encounter within the health system.

We frame disagreement between the claims-based algorithms and a patient's designated PCP of record as a miss-classification. Based on these data, we cannot discern whether these cases were in fact "misses" or whether the patient intentionally sought out more than one PCP. From the perspective of the health care system, the designated PCP is held accountable for the patient's health care utilization and adherence to quality of care measures. Patients are also asked to confirm the name of the designated PCP at every encounter with the health system. In health services studies, there is an assumption that there is a single, responsible PCP. However, in some cases, there may not be any responsible PCP or there may be more than one. To the extent the majority and plurality claims-based algorithms are used by researchers to identify a patient's likely PCP, we find in this sample that these algorithms are often right, however, there is some variation by patient characteristics.

Several other study limitations should be noted. These data rely on electronic health record data entries of the designated primary care physician. In this academic health system, the PCP of record is confirmed by the patient at each visit, but can be assigned by the health care system in cases of new patient enrollment or physician exit. In this health system, non-physicians such as nurse practitioners and physician assistants do not serve as the PCP of record. While the health system uses this field to identify the accountable PCP and patients are asked to confirm the name of their PCP at each visit, it is possible that patient's may not view the PCP of record to be their true primary physician. Our confidence in the validity of the PCP of record indicator substantially increases among patients who receive 100% of their care within the health care system. In this patient subgroup we find that the plurality algorithm using primary care visits identifies the PCP of record in nearly 90% of cases. This study draws on a sample of older adults with diabetes who are medically housed at a Midwestern academic health system that includes primary, specialty, and inpatient services. These results may not generalize to other geographic areas or disease groups.

Our findings suggest that there are differences in the accuracy of claims-based PCP assignment algorithms compared to EHR records of the PCP of record, and variation by

patient-level characteristics. Researchers should approach these algorithms with caution, and when possible use a tie breaker to improve the accuracy of their approach.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Haggerty JL, Reid RJ, Freeman GK, et al. Continuity of care: a multidisciplinary review. *Bmj*. 2003; 327:1219–1221. [PubMed: 14630762]
2. Burgers JS, Voerman GE, Grol R, et al. Quality and Coordination of Care for Patients With Multiple Conditions: Results From an International Survey of Patient Experience. *Evaluation & the Health Professions*. 2010; 33:343–364. [PubMed: 20801976]
3. Pham HH, Schrag D, O'Malley AS, et al. Care patterns in Medicare and their implications for pay for performance. *N Engl J Med*. 2007; 356:1130–1139. [PubMed: 17360991]
4. Weir DL, McAlister FA, Majumdar SR, et al. The Interplay Between Continuity of Care, Multimorbidity, and Adverse Events in Patients With Diabetes. *Med Care*. 2016
5. Hussey PS, Schneider EC, Rudin RS, et al. Continuity and the costs of care for chronic disease. *JAMA internal medicine*. 2014; 174:742–748. [PubMed: 24638880]
6. Reid RJ, Fishman PA, Yu O, et al. Patient-centered medical home demonstration: a prospective, quasi-experimental, before and after evaluation. *Am J Manag Care*. 2009; 15:e71–87. [PubMed: 19728768]
7. Van Walraven C, Oake N, Jennings A, et al. The association between continuity of care and outcomes: a systematic and critical review. *Journal of Evaluation in Clinical Practice*. 2010; 16:947–956. [PubMed: 20553366]
8. Shah BR, Hux JE, Laupacis A, et al. Administrative data algorithms can describe ambulatory physician utilization. *Health Serv Res*. 2007; 42:1783–1796. [PubMed: 17610448]
9. Mehrotra A, Adams JL, Thomas JW, et al. The effect of different attribution rules on individual physician cost profiles. *Ann Intern Med*. 2010; 152:649–654. [PubMed: 20479030]
10. Segal JB, DuGoff EH. Building blocks for measuring care coordination with claims data. *Popul Health Manag*. 2014; 17:247–252. [PubMed: 24606582]
11. Nyweide DJ, Bynum JP. Relationship Between Continuity of Ambulatory Care and Risk of Emergency Department Episodes Among Older Adults. *Ann Emerg Med*. 2016
12. DuGoff EH, Banteen-Roche K, Anderson GF. Relationship between continuity of care and adverse outcomes varies by number of chronic conditions among older adults with diabetes. *Journal of Comorbidity*. 2016; 6:65–72. [PubMed: 29090176]
13. Nyweide DJ, Anthony DL, Bynum JP, et al. Continuity of care and the risk of preventable hospitalization in older adults. *JAMA internal medicine*. 2013; 173:1879–1885. [PubMed: 24043127]
14. Pollack CE, Hussey PS, Rudin RS, et al. Measuring Care Continuity: A Comparison of Claims-based Methods. *Med Care*. 2013
15. Bynum JP, Bernal-Delgado E, Gottlieb D, et al. Assigning ambulatory patients and their physicians to hospitals: a method for obtaining population-based provider performance measurements. *Health Serv Res*. 2007; 42:45–62. [PubMed: 17355581]
16. Thorpe CT, Flood GE, Kraft SA, et al. Effect of patient selection method on provider group performance estimates. *Med Care*. 2011; 49:780–785. [PubMed: 21617570]
17. Niefeld MR, Braunstein JB, Wu AW, et al. Preventable hospitalization among elderly Medicare beneficiaries with type 2 diabetes. *Diabetes Care*. 2003; 26:1344–1349. [PubMed: 12716786]
18. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care*. 1998; 36:8–27. [PubMed: 9431328]
19. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav*. 1995; 36:1–10. [PubMed: 7738325]

20. Aday LA, Andersen R. A framework for the study of access to medical care. *Health Serv Res.* 1974; 9:208–220. [PubMed: 4436074]
21. Reddy A, Pollack CE, Asch DA, et al. The Effect of Primary Care Provider Turnover on Patient Experience of Care and Ambulatory Quality of Care. *JAMA internal medicine.* 2015; 175:1157–1162. [PubMed: 25985320]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Patient and Provider Characteristics in 2008

	2008
Patient Characteristics	
N	3,658
Patient Years	15,624
Age (Mean [SD])	75.9 (6.9)
Age Group (%)	
65–74	46.7
75–84	40.3
85+	13.1
Female (%)	54.7
Medicaid (%)	9.2
White (%)	95.7
Number of Chronic Conditions (Mean [SD])	3.6 (2.4)
3 or more chronic conditions (%)	60.0
Physician Visits (%)	
5	34.1
6 to 10	34.3
>10	31.5
Provider Characteristics	
N	5,261
N Primary Care Physicians	1,316
N Designated Primary Care Physicians	505
Average Panel Size	10.6 (13.5)

Note: The number of primary care physicians is calculated as the number of physicians seen by patients for primary care visits during the entire sample period. A designated primary care physician is a physician selected by a patient and recorded in the electronic health record system.

Table 2

Concordance between Claims-Based Algorithms and Designated Primary Care Physician by Type of Physician Visit

	All Visits			Primary Care Physician Visits		
	Majority Match %	Plurality Match %	P-Value	Majority Match %	Plurality Match %	P-Value
Overall	25.4	56.2	<0.001	78.0	83.6	<0.001
Tie Breaker						
Amount billed	63.3	63.2	0.326	84.8	85.1	<0.001
Last Visit	45.3	60.2	<0.001	82.9	85.3	<0.001
Longest Timespan	59.8	62.0	<0.001	85.9	85.8	0.374

Note. The majority physician designates the physician with whom the patient had greater than 50% of visits. The plurality rule designates the physician with whom the patient had the most visits in a year. P-values for McNemar's test are reported in parentheses comparing the majority and plurality algorithms.

Table 3

Concordance between Claims-Based Algorithms and Designated Primary Care Physician by Patient Characteristics Using No Tie Breaker (Percent Correct)

	All Visits			Primary Care Physician Visits		
	Majority Match %	Plurality Match %	P-Value	Majority Match %	Plurality Match %	P-Value
Race						
White	25.4	56.3	<0.001	78.5	83.9	<0.001
Not White	23.5	53.0	<0.001	69.1	77.0	<0.001
Medicaid						
No	25.5	56.5	<0.001	79.1	84.5	<0.001
Any	23.9	52.9	<0.001	68.0	75.1	<0.001
Chronic Conditions						
2	26.8	55.5	<0.001	77.8	83.0	<0.001
3+	24.5	56.6	<0.001	78.2	83.8	<0.001
% Within System Visits						
100%	29.6	57.6	<0.001	81.8	88.0	<0.001
<100%	24.6	55.9	<0.001	77.4	82.8	<0.001
PCP Change						
No	27.9	60.0	<0.001	60.0	89.8	<0.001
Yes	23.3	52.5	<0.001	70.8	76.8	<0.001
Physician Visits						
5	45.4	53.6	<0.001	82.0	83.5	<0.001
6 to 10	27.4	57.2	<0.001	75.8	82.3	<0.001
>10	13.4	52.4	<0.001	69.5	77.8	<0.001

Note. The majority physician designates the physician with whom the patient had greater than 50% of visits. The plurality rule designates the physician with whom the patient had the most visits in a year. P-values for McNemar's test are reported in parentheses comparing the majority and plurality algorithms.