Comparison of Electronic Health Record–Based and Claims-Based Diabetes Care Quality Measures: Causes of Discrepancies

Michael Barton Laws iD *, Joanne Michaud, Renee Shield, William McQuade, and Ira B. Wilson*

Objective. To investigate magnitude and sources of discrepancy in quality metrics using claims versus electronic health record (EHR) data.

Study Design. Assessment of proportions of HbA1c and LDL testing for people ascertained as diabetic from the respective sources. Qualitative interviews and review of EHRs of discrepant cases.

Data Collection/Extraction. Claims submitted to Rhode Island Medicaid by three practice sites in 2013; program-coded EHR extraction; manual review of selected EHRs.

Principal Findings. Of 21,030 adult Medicaid beneficiaries attributed to a primary care patient at a site by claims or EHR data, concordance on assignment ranged from 0.30 to 0.41. Of patients with concordant assignment, the ratio of patients ascertained as diabetic by EHR versus claims ranged from 1.06 to 1.14. For patients with concordant assignment and diagnosis, the ratio based on EHR versus claims ranged from 1.08 to 18.34 for HbA1c testing, and from 1.29 to 14.18 for lipid testing. Manual record review of 264 patients discrepant on diagnosis or testing identified problems such as misuse of ICD-9 codes, failure to submit claims, and others.

Conclusions. Claims data underestimate performance on these metrics compared to EHR documentation, by varying amounts. Use of claims data for these metrics is problematic.

Key Words. Quality measurement, EHRs, claims data

Efforts to improve health care quality through process and/or outcome indicators date to the development of the Healthcare Effectiveness Data and Information Set (HEDIS) in 1991 (Mainous and Talbert 1998). As claims data are already collected and available, they were widely used (Lohr 1990). Early

investigators found substantial discrepancies between measures derived from medical records and claims (Jollis et al. 1993). As electronic health records (EHRs) became more widely used and capable, the prospect that automated data extraction from EHRs could produce accurate quality metrics became enticing.

Quality metrics must evaluate several variables with acceptable accuracy. For the denominator, these include the following: (1) proper assignment of responsibility for patients to a given provider institution or site and accurate assessment of the duration of assignment; (2) proper assignment of patients to a relevant attribute, such as a diagnosis or age range, accounting for exclusion criteria; and (3) for some purposes, proper assignment of patients to a payer. Ascertainment of the numerator, such as receipt of a procedure, must be possible for all patients in the denominator.

We are not aware of any published studies that evaluate accuracy of assignment of patients to providers using claims data. Published studies have found varying degrees of discrepancy between automated and manual extraction of data from EHRs. A 2007 study that assessed quality metrics for management of heart failure found good agreement for most measures, but automated extraction failed to identify accepted contraindications for prescription of Warfarin (Baker et al. 2007). Failure to identify contraindications for prescribing was found in another study based on data from 2009 to 2010 (Danford et al. 2013). A study of Veterans Administration (VA) clinics in 1999 found that blood pressure data were present in a structured field for 60 percent of visits, but they were found in an additional 11 percent in visit notes (Borzecki et al. 2004). Another VA study found variable agreement between manual review and automated data extraction depending on the quality metric. Providers do not always transfer information from notes and scanned documents into structured fields even when they have received training in this regard (Parsons et al. 2012), but otherwise information about why these discrepancies occur is sparse. Generally, automated data extraction underestimated providers' performance (Urech et al. 2015).

Address correspondence to Michael Barton Laws, Ph.D., Department of Health Services, Policy and Practice, Brown University, G-S121-7, Providence, RI 02912; e-mail: michael_barton_laws@ brown.edu. Joanne Michaud, M.S., Renee Shield, Ph.D., and Ira B. Wilson, M.D., M.Sc., are with Department of Health Services, Policy and Practice, Brown University School of Public Health, Providence, RI. William McQuade, D.Sc., M.P.H., is with the Center for Primary Care and Prevention, Warren Alpert Medical School, and Rhode Island Executive Office of Health and Human Services, Warwick, RI.

A comparison of claims data with automated extraction from EHRs for preventive care quality metrics found considerable variation depending on the service, with EHRs generally more sensitive (Heintzman et al. 2014). A study found substantial underascertainment of services by claims as compared to EHR data for four services (DeVoe et al. 2011). Another found substantial disagreement on metrics for children's preventive services (Angier et al. 2014). A study of patients with diabetes in one outpatient clinic, and insured by an HMO, found 92.7 percent receiving an HbA1c test in 1998 using record review versus 36.3 percent identified through claims (Maclean et al. 2002). Another study found that electronic data extraction had 97.6 percent sensitivity for detecting diabetes, but claims data had only 75 percent sensitivity (Tang et al. 2007).

The American Recovery and Reinvestment Act of 2009 accelerated adoption of EHRs (Blumenthal 1993, 2009, 2010), and the Patient Protection and Affordable Care Act (ACA) (Centers for Medicare and Medicaid Services, 2016) mandated regulations to standardize EHRs to improve health care quality through Meaningful Use certification and associated financial incentives (Office of the National Coordinator for Health Information Technology 2014; Centers for Medicare and Medicaid Services 2017).

As both claims and EHR data have limitations for measuring care quality (DeShazo and Hoffman 2015; McCoy et al. 2016), the primary goal of this project was to determine whether Medicaid pay-for-performance incentive payments could be based on data extracted from structured fields in an EHR, as opposed to claims data. Our novel approach compared denominators from three variables: patient enrollment in Medicaid; assignment of responsibility for the patient to one primary care practice; and diagnosis with diabetes. To explain sources of discordance, we used a mixed methods strategy, including qualitative data.

METHODS

Data Use Agreements/IRB Review

The protocol was specified by Data Use Agreements between Brown University and RI Executive Office of Health and Human Services, and between Brown University and the clinical sites. The project was approved by the Brown University Institutional Review Board (IRB) and the IRB of one site that required local review.

Quality Measures

We chose annual hemoglobin A1c (HbA1c) testing and low-density lipoprotein cholesterol (LDL-C) testing in adult patients with a diagnosis of diabetes mellitus (DM) as our measures. Both tests are billable and should result in a claim to document their occurrence. These are included in the 15 EOHHS measures and are sufficiently common to ensure adequate sample size. Both measures are approved by NCQA for assessment using administrative data for managed care accreditation.

Clinical Sites

We selected three large primary care practices in RI based on familiarity with quality reporting, proficiency at EHR data extraction, size and diversity of populations served, and use of different EHRs. See Table 1 for self-reported attributes of each site. Rhode Island (RI) Patient-Centered Medical Homes and Federally Qualified Health Centers submit EHR quality measures to the Rhode Island Quality Institute (RIQI) in connection with previous initiatives (Care Transformation Chronic Care Sustainability Initiative 2017; Rhode Island Quality Institute 2017). The selected facilities had developed algorithms for reporting this information, with technical assistance from RIQI. As this was a pragmatic study, we asked them to continue to use the same extraction algorithms.

Identifying Medicaid Patients

Sites. Analysts at each clinical site identified Medicaid patients 18–64 years of age who received care at that site during calendar year 2013. At Site 1, the analyst selected patients seen in 2013 who were enrolled in Medicaid (including Medicaid managed care [MC] programs operated by a private insurer) as their primary insurer at the time of their last visit in 2013. At Sites 2 and 3, the analyst selected patients seen in 2013 with any charge billed to Medicaid.

Medicaid. We used the RI Medicaid enrollment files to identify members enrolled in MC plans in 2013. We selected MC members whose assigned PCP matched a PCP identifier included on lists received from the sites.

We selected FFS Medicaid members from state administrative files and attributed members to providers using an adaptation of the Medicare Multi-Payer Advanced Primary Care Practice (MAPCP) Demonstration Assignment

Table 1:   Attributes of Clinical Sites and Their Patients, 2013

Site	1	2	3
Year founded	1974	1990	1999
Year designated a PCMH	2008	2011	2008
EHR/year adopted	eClinicalWorks/2006	NextGen/2007	Centricity EHR/1999
EHR functionality	Front- and back-end connection to billing data	Front- and back-end connection to billing data	No connection to billing data
Patients served in 2013	41,828	13,332	7,892
Medical visits in 2013	120,769	38,560	24,725
% patients below poverty level, 2013	49.16	30.13	Not available
% patients nonwhite, 2013	14 (19% unreported)	32 (19% unreported)	25
Patient ethnicity (% Hispanic), 2013	Hispanic/Latino = 14% (3.5% unreported)	Hispanic/Latino = 54% (6.17% unreported)	Hispanic/ Latino = 21%
% patients enrolled in Medicaid (excluding dual enrollment), 2013	38	48	39

Algorithm. Members were attributed to a provider when there were more visits to that provider than to any other in 2012–2013. When the number of visits to two or more providers was equal, we attributed the member to the provider with the most recent visit. We assigned members whose attributed PCP matched a PCP identifier received from a study site. Finally, we selected the FFS members who were continuously enrolled in 2013, with no more than one gap of 45 or fewer days.

Identifying Diabetic Patients

Sites. For all patients selected by the clinical sites, we received patient identifiers (full name, date of birth, SSN), DM status in 2012, DM status in 2013, and 2013 HbA1c and LDL-C test dates and results for DM patients. Data came from structured fields only. Because Sites 1 and 2's EHRs connected to their billing systems, they ascertained DM status from both the EHR problem list and the billing module, which includes diagnosis codes for visits.

Site 1 flagged DM patients from problem list content from previously built year-end quality measure tables, based on an International Classification of Diseases ver. 9 (ICD9) diagnosis code for DM (250.x, 357.2, 362.0x,

366.41, or 648.0); or an ICD9 diagnosis code for DM from the billing data for at least one visit in 2012 or 2013.

Site 2 mined its problem lists to flag DM patients, searching for text of a diagnosis of diabetes with onset date = <2013. (The patient problems table on the back end does not have an ICD9 code column.) The search eliminated items such as "Family history of diabetes mellitus" or "Prediabetes." DM assessment fields were applied as for Site 1 above.

Site 3 coded for a DM diagnosis if there was an ICD9 diagnosis code of 250.x on the problem list at the end of 2012 or 2013.

Medicaid. We constructed analytic files for claims for all encounters, regardless of site of care, during 2012 and 2013, including patient identifiers, dates of service, ICD9 billing codes, and Current Procedural Terminology (CPT) billing codes for each date of service. We used the HEDIS criteria for identifying denominator patients: (1) dispensed insulin or oral hypoglycemics/antihyperglycemics during 2012 or 2013 on an ambulatory basis; or (2) had at least two outpatient visits, or one encounter in an acute inpatient or Emergency Department setting, with a diagnosis of DM in 2012 or 2013.

Ascertaining Testing

Sites. The sites extracted indicators of test performance in calendar 2013 from the relevant structured fields in their EHR laboratory modules.

Medicaid. For claims data, we coded for an HbA1c test if there was at least one HbA1c CPT billing code (83036, 83037, 3044F, 3045F, or 3046F) in 2013; and for an LDL-C test if there was at least one LDL-C CPT billing code (3048F, 3049F, 3050F, 80061, 83700, 83701, 83704, or 83721) in 2013.

Analysis Steps

Analyses used SAS, version 9.4 (SAS Institute, Cary, NC); Stata, version 14.0 (StataCorp, College Station, TX); and Excel, version 2010 (Microsoft, Redmond, WA). We matched data from EHRs and claims using SSNs, and built a dataset of all fields for all persons included from either source.

Qualitative Data

MBL and RS visited each site on two occasions to interview and shadow providers and other staff. To review electronic health records, MBL used a

preselected list of patients with discordance between claims and EHR data on either DM diagnosis or receipt of a test. Patient identifying information was linked to the data collection instrument by a sequential number. Following data collection, the identifying file was destroyed. The intention was to include only patients assigned to the site by both the site and claims data. See Table 2 for sampling criteria. This sample consisted of 264 patients: 111 Site 1 patients, 63 Site 2 patients, and 86 Site 3 patients.

The reviewer was blinded to the reason for inclusion in the sample. He viewed EHR and claims data, if available, recorded the following data for all patients in the sample using a relational database application, and took notes describing instances of missing or anomalous data, and any apparent explanations.

- 2012/2013 problem list entries (DM diagnosis y/n, DM ICD9 code, DM start date, and DM end date, if any)
- 2012/2013 clinic visits (visit date, payer if available, DM y/n, billing DM ICD9 code)
- 2013 LDL-C tests (date, value, location, billed y/n if on-site and payer if available, not in structured field y/n), and date entered if not in structured field.)
- 2013 HbA1c tests (date, value, location, billed y/n if on-site and payer if available, not in structured field y/n, and date entered if not in structured field)

Table 2:   Sampling Strategy for Manual Record Review

Discordance	Sample Criteria	Data Collected
Discordant for DM Dx	All patients ascertained w/ DM by claims but not EHR Patients ascertained w/ DM by EHR but not claims randomly sampled to yield approximately equal number as above	Problem list entries: DM diagnosis y/n; DM ICD9 code; DM start date; DM end date if any 2012/2013 clinic visit dates; payer if available; DM in billing Dx code y/n; DM ICD9 billing code
Discordant for HbA1c or LDL-C test	Drawing from patients concordant for both site assignment and DM Dx, all patients who ascertained as receiving a test in 2013 by claims but not EHR Patients ascertained as receiving a test in 2013 by EHR but not claims randomly sampled to yield approximately equal number as above	2013 LDL-C tests: date; value; performer; billed if on-site; payer if available; in structured field y/n 2013 HbA1c tests: Date; value; performer; billed if on-site; payer if available; in structured field y/n

As reviewer was blinded to reasons for inclusion in the sample, he collected all data for all patients.

At Site 3, the reviewer determined whether an ICD9 code for DM was listed in the visit summary, which was presumably the basis for any claim; he could not see the actual claim. Patients included in the sample by virtue of DM diagnosis discordance were flagged for 2-year lookback (i.e., 2012 and 2013); those included because of an HbA1c or LDL test discordance were flagged for 1-year lookback (2013).

RESULTS—QUANTITATIVE STUDY

Concordance on Assignment (Attribution)

Table 3 shows the comparison of persons assigned to each site by the Medicaid algorithm to those reported by the site.

For example, row 1 of Table 3 shows that of the patients assigned by Medicaid to Site 1, Site 1 identified 51 percent as their patients. Row 2 shows that of the patients reported by Site 1 to be Medicaid patients for whom they provided care, 67 percent were also assigned to Site 1 by Medicaid. The Sites and Medicaid agreed on assignment in 41 percent, 38 percent, and 30 percent of cases, respectively (row 3). (See Figure 1 for a Venn diagram.)

Because concordance on assignment was so poor, we report further results only for patients assigned to a given site by both the site and Medicaid. Discrepancies on diagnosis and testing were greater in all cases when including patients discordant on assignment.

Diabetes Diagnosis

For persons with concordant assignment to site, comparison of DM diagnosis according to claims or EHR data is shown in Table 4.

For example, for Site 1, the claims data indicate a 10 percent prevalence of DM diagnosis, while EHR data indicate an 11 percent prevalence. The last

Table 3:   Concordance on Assignment of Patient to Provider

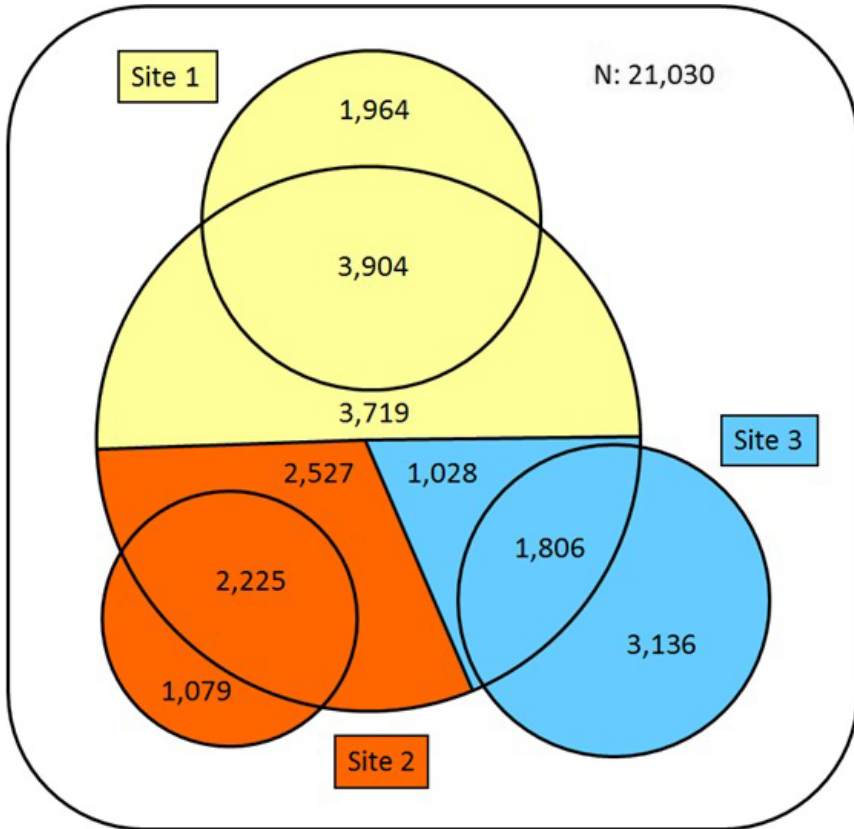| Site	1	2	3
Assignment rate*	0.51	0.47	0.64
Panel rate[†]	0.67	0.67	0.37
Concordant rate[‡]	0.41	0.38	0.30

*Reported by EHR given assignment by claims.
[†]Assigned by claims given reported by EHR.
[‡]Reported by both claims and EHR given inclusion on either list.

Figure 1:   Concordance/Discordance on Assignment to Site [Color figure can be viewed at wileyonlinelibrary.com]

Notes. Numbers of persons included in Medicaid claims dataset (large circle) and clinical EHR datasets (three smaller circles). Attributions of persons from claims data and inclusion of persons in EHR data are color-coded by clinical site. The sum of individual segments (21,388) is larger than the total number of unique persons in the combined datasets (21,030), due to multiple attributions or multiple clinics visited.

row of Table 4 indicates that the EHR shows 6 percent, 11 percent, and 14 percent more DM cases than claims, respectively.

HbA1c Testing

Table 5 shows the proportion of persons identified as diabetic by the respective data source who received an HbA1c test in 2013 according to the same source.

Table 4: For Persons Concordant on Assignment to Site, Proportions with DM Diagnosis

Parameter	Site 1	Site 2	Site 3
Claims*	0.10	0.08	0.09
EHR[†]	0.11	0.09	0.11
EHR/Claims proportion	1.06	1.11	1.14

*Proportion of patients with a claim for dispensed insulin or oral hypoglycemics/antihyper-glycemics during 2012 or 2013 on an ambulatory basis; or had at least two outpatient visits, or one encounter in an acute inpatient or Emergency Department setting, with a diagnosis of DM in 2012 or 2013.
[†]As identified by protocol specific to the site.

Table 5: Prevalence of HbA1c and LDL-C Testing in 2013 for DM Patients with Concordant Assignment

Parameter	Site 1	Site 2	Site 3
HbA1c			
Claims*	0.89	0.82	0.04
EHR[†]	0.96	0.94	0.65
EHR/Claims	1.08	1.14	18.34
LDL-C			
Claims*	0.63	0.37	0.04
EHR[†]	0.82	0.85	0.50
EHR/Claims	1.29	2.34	14.18

*A paid claim for the test from any performer in 2013.
[†]Presence of a test result in 2013 in the laboratory module.

For example, for persons attributed to Site 1 by claims *and* on the Site 1 list, 89 percent of the patients identified as diabetic had an HbA1c test in 2013 according to claims data, and 96 percent had an HbA1c test according to Site 1. Site 3 had a lower proportion of HbA1c testing according to EHR data, but a dramatically lower proportion according to claims (0.04 vs. 0.89 and 0.82).

LDL Testing

For those persons with concordant assignment, Table 5 compares the proportion of LDL testing for those identified as diabetic by the respective data source. For example, for Site 1, 63 percent of patients identified as diabetic had an LDL test according to claims data, and 82 percent had a test according to EHR data. LDL testing proportions were more discrepant than the HbA1c rates in Sites 1 and 2, with the same dramatic underreporting of LDL testing by claims in Site 3 seen for HbA1c testing.

RESULTS—QUALITATIVE STUDY

Blood Testing and Recording of Results

Sites 1 and 2 had on-site rapid testing for HbA1c and lipid panels; and EHRs with front- and back-end connection to the billing system. Both also had colocated unaffiliated laboratories with electronic connections to the EHR. At Sites 1 and 2, the medical assistant (MA) assigned to the provider saw the patient prior to a diabetes follow-up visit, collected finger stick blood samples, and transported them to the laboratory for automated analysis. The MA then manually entered the test results into the EHR. Test performance was automatically transmitted to the billing system and included in the billing summary for the visit.

Tests performed at the colocated independent laboratories were electronically transmitted to the EHR. Results of tests performed elsewhere would come as a faxed or mailed document, which a nurse would scan and append to the EHR. The medical record departments were charged with entering these data into the appropriate structured field, but informants said this did not always occur.

Site 3 did not have rapid testing, but the hospital's laboratory on the campus connected electronically to the EHR. The provider would print out the test order at the time of the visit and instruct the patient to have it performed. Most patients used the hospital laboratory, but there was no clerical staff to do data entry for patients tested elsewhere. We were told that nurses did this task inconsistently.

Visit Claims

Providers at all sites entered the billing diagnosis codes for the visit along with the visit notes. At Sites 1 and 2, these were electronically transmitted to the billing module. At Site 3, the visit note was a structured electronic document which was printed and transported in paper form to a billing office, where the claim was created manually.

Problem Lists

There were standard criteria at Sites 1 and 2 for entering a diagnosis of DM into the problem list, based on HbA1c or fasting glucose. Individual providers may decide whether to remove DM from the problem list if patients remit

(e.g., due to bariatric surgery or radical lifestyle changes). At Site 3 (see below), the problem list was less structured.

Specialty Services

At Site 3, the hospital has a full suite of specialty services. Patients of the primary care clinic may receive diabetes care from an endocrinologist.

Medical Records: General Observations

Due to the architecture of the various EHRs, the reviewer inevitably saw visit notes at Sites 1 and 3, as this was the only way to access billing information. He looked at visit notes at Site 2 when necessary to resolve a question.

Due to software upgrades, at Sites 1 and 2 in most cases, the onset date for DM in the problem list could not be definitively determined. At Sites 1 and 2, where the reviewer had access to claims records, diagnostic codes for visit claims and in-house laboratories could be determined.

Medical Record Review

Assignment to Site. Although the sample was intended to include only patients assigned to the site by both Medicaid and the site, we could not locate an EHR for one or two patients at all three sites (perhaps due to death or changed identifying information since 2013). Some additional patients were not seen during 2012–2013, although they may have been seen in 2011 and/or 2014. At Site 3, the EHR included letters to two patients informing them they had been terminated due to repeated missed appointments.

Some patients received prenatal care at Sites 1 and 2, but primary care elsewhere. For MC patients, this would result in a discrepancy of assignment. Reimbursement for the entirety of prenatal care was typically through bundled payment. However, there were individual claims for visits in which an additional billable procedure was performed. This could have resulted in the sites having the most visits by a patient in the year, leading, in the case of a FFS patient, to assignment of the patient to the site by claims. The EHR extraction algorithms at the sites would also have identified these patients, resulting in false concordant assignment to the site. At Site 3 when primary care patients' diabetes was managed by an endocrinologist, laboratory results were sometimes observed in visit notes, but not in the laboratory module.

Insurance Status. Noncontinuous enrollment at Sites 1 and 2 was common. Many did not have Medicaid until 2013, sometimes late in the year. A common pattern was for people to acquire Medicaid coverage for 1 or 2 months, lose it for 1 or 2 months, and then regain it. This appears to be due to presumptive eligibility, which allows the provider to bill prior to final determination of eligibility, but may result in an interruption of coverage during determination. At Site 1, a few patients were enrolled in the RiteShare program, in which Medicaid pays the premium for private insurance. The site may have regarded these patients as Medicaid patients, but Medicaid would not have seen their claims.

Diagnosis. Some prenatal patients at Sites 1 or 2 may have been erroneously designated as diabetic because a visit containing a billable procedure had gestational diabetes miscoded either as 250.x or as 648.6 (diabetes complicating pregnancy). The most common billable procedure associated with prenatal care was a fetal nonstress test.

Other patients at all three sites with extensive comorbidity and numerous visits did not have any version of DM on the problem list; had no visits with a DM diagnosis during the study period; and had no indication of prediabetes or impaired fasting glucose, or a prescription for metformin. Their ascertainment with DM may have resulted from an encounter elsewhere.

One patient at Site 1 was diagnosed with DM in early 2013, but the diagnosis was not mentioned again and was not added to the problem list until 2014, when DM management began. A few patients were not seen again after the diagnosis of DM.

Remission of DM and removal from the problem list occasionally occurred. Some patients were never diagnosed with diabetes, but only with prediabetes. Patients diagnosed with prediabetes may be prescribed metformin, an ascertainment criterion for claims data but not for the EHR. Prediabetes may also be miscoded in a visit note as 250.x. At Site 3, the problem list allowed free-text entries in association with ICD9 codes. Providers would occasionally write "family history of diabetes" or "risk of diabetes," in association with the ICD9 code for DM. They seemed to be using the problem list as a reminder of concerns, not necessarily a record of actual diagnoses. This would have resulted in erroneous ascertainment in the EHR extraction.

At Site 3, providers would also occasionally enter an ICD9 code of 250.x in the visit summary, when diabetes was discussed as a concern, but not diagnosed. In one case, a patient received a diagnosis of diabetes in one visit,

was counseled about diet and exercise, and later told that it had remitted. The diagnosis was never entered in the problem list.

Testing. At Sites 1 and 2, in-house HbA1c or lipid analyzer tests carried out in connection with diabetes follow-up visits were billed with few exceptions. At Site 1, in-house laboratories were occasionally performed outside a billed visit —for example, in a visit for group diabetes education. Claims were not normally generated for these. At Site 2, one test was observed to have been billed, but it was not in the laboratory module.

   At Sites 1 and 2, a few patients were dual Medicare and Medicaid eligible at some time during the period. In these cases, Medicare receives the claim for procedures, while Medicaid may only see the claim for the visit co-pay.

   For two visits of separate patients at Site 1, there were no corresponding claims, evidently due to failure of processing. At Site 2, one claim for a visit was not paid because it was billed to the patient's dental insurance. This visit included both tests. This could have resulted in discrepancy on testing; that is, the test would be in the EHR but not in the claims data. It could also have resulted in a discrepancy in diagnosis, as the visit with a DM diagnosis code would not have been paid but would be in the EHR.

   A few tests recorded in the EHRs at Sites 1 and 2 were carried out in connection with hospitalization or ED visits. These may not have generated a claim. At Site 3, tests by other performers were not found. Finally, the lipid analyzer sometimes failed to calculate the LDL due to high triglycerides or an insufficient amount of blood. Both EHR and claims data would assess these tests as performed, as the fact of performance is in the EHR, and these tests were billed.

DISCUSSION

Three points are clear from quantitative results. First, assignment of patients to a primary care site and tracking duration of assignment are major sources of variance. Second, discrepancies in ascertainment of diagnosis between claims data and EHRs are fairly small when considering only cases with concordant assignment. Third, while claims data underestimated performance, proportions of HbA1c testing were within 8–14 percent of EHRs in two sites. These results suggest that in sites where claims processing is tightly monitored, and proper assignment to the site can be established, claims-based rates could have some validity for site attestation of HbA1c testing. However, our results do

not support acceptance of lipid testing using claims data as a valid estimate of performance.

The large discrepancies in assignment to a primary care provider between the claims algorithm and the sites' EHR extractions present a challenge for quality measurement. The largest contributor to discordance was managed care patients, who do not necessarily see the provider to whom they are formally assigned. This could be addressed by a change in the insurer's practice for attributing managed care patients to a primary care provider. Misattribution of fee-for-service patients to a PCP could be ameliorated by a more thorough algorithm that weeds out patients who have multiple specialty visits to a facility that is not their PCP. However, we do not see any obvious bias introduced by limiting the analysis to patients with concordant assignment.

In general, the frequency of recorded events will be lower in claims data because additional processing steps are needed. While providers and laboratories have an incentive to submit claims whenever they can, claims processing is routine, low-paid work prone to occasional error. The average reimbursement for an HbA1c test of about $10, when added to the incentive associated with meeting performance standards, should be enough incentive to submit all the HbA1c claims. However, this did not seem to be the case when, for example, a rapid test was performed during an otherwise unbilled nursing visit. Site 3 did not appear to bill for these tests.

While EHR extraction is more accurate, it is far from perfect. Regardless of the EHR software used, humans perform the often tedious entry of data. Providers do not have any particular incentive to be diligent and consistent about entering ICD9 codes, unless management policy creates one. Many codes are associated with DM, not all of which represent actual diagnosis of the condition. As the 250.x code for DM is encountered most frequently, providers may use it from habit, rather than take the trouble to look up the precise code. They also may not use the ICD9 code to indicate diagnosis, but as a reminder of potential risk or to classify a concern addressed in a visit. Finally, practice sites have difficulty replicating NCQA criteria for continuous enrollment.

The sites were chosen for their high rate of patients enrolled in Medicaid and the sophistication of their EHRs. If these relatively large and sophisticated organizations make these errors, smaller practices with fewer resources may make more errors more often.

These issues would be less problematic if the discrepancies were consistent in size, but they were not. While no simple adjustment to claims-based

measures appears possible from these results, it is possible for sites to track their HbA1c submissions to be sure that their payers are being properly notified of the services they are providing. Practice sites can (and perhaps should) be alerted to the importance of making sure that all their services are being properly billed.

Quality measurement and related pay-for-performance (P4P) incentives should probably be based on data from EHRs, not from claims. This recommendation is tempered by the extent to which human factors can drive what data end up in structured fields and thus become part of electronic data pulls. Given how different EHRs are from each other, specific EHR modules that practices can afford to purchase, and the difficulty of designing and managing procedures by which humans interact with EHRs, it is not clear that quality data extracted from EHRs can validly compare practices with one another. While there are efforts to certify EHRs for this purpose (Centers for Medicare and Medicaid Services 2017), these reformed certification processes may not address many of the issues that we identified in this study.

For our sample of practices, reliable process measurement, whether using EHR or claims data, would require investment in training and management procedures, and probably ongoing auditing. Not all practices have the resources to make this investment, and available incentives may not be sufficient to motivate them. Payers could provide training and technical assistance, and invest in auditing. While there may well be a long-term social benefit to such investment, the individual payer may not expect that patients will remain their responsibility for the long term. Medicaid programs may view this problem differently than commercial insurers.

Some argue that quality measurement and improvement should focus on outcomes, not processes (Porter, Larsson, and Lee 2016). While outcome measurement poses different problems than process measurement (most important, the challenge of risk adjustment), as this study shows, valid process measurement is far more complex than many might suspect. The time, energy, and resources spent on process measurement might be better applied toward getting outcome measurement right.

This study has several limitations. We studied one Medicaid program at a particular time in its history, and only three practices, all large and experienced with EHRs. Studying different practices at different times in different states might have produced dissimilar findings. Findings for payers other than Medicaid may also have been different. Our goal was to study practices as they conducted routine business, not to understand what levels of concordance might be under optimal, and rarely obtainable, circumstances. We

suspect that our findings are generalizable to many practices in many states where providers are struggling to implement and "meaningfully" use EHRs.

This study also had notable strengths. The qualitative work is highly informative. Future qualitative inquiry on this subject could include interviews of data entry personnel and different levels of providers. Observations in conjunction with interviews could shed light on real-time processes of clinical testing and recording, and claims entry. In addition, we purposefully sampled sites with different EHRs, which allowed us to characterize the wide variety of challenges each of these EHRs creates, and how different they are from one other.

In conclusion, if claims-based systems are used to document or verify practice site attestation of performance, care must be taken to assure complete and accurate submission and payment of relevant claims. In addition, we documented site and EHR-specific variations in the ways in which humans interact with EHR technologies. These variations may call into question the validity of cross-practice comparisons.

ACKNOWLEDGMENTS

REFERENCES

Angier, H., R. Gold, C. Gallia, A. Casciato, C. J. Tillotson, M. Marino, R. Mangione-Smith, and J. E. DeVoe. 2014. "Variation in Outcomes of Quality Measurement by Data Source." *Pediatrics* 133 (6): e1676–82.

Baker, D. W., S. D. Persell, J. A. Thompson, N. S. Soman, K. M. Burgner, D. Liss, and K. S. Kmetik. 2007. "Automated Review of Electronic Health Records to Assess Quality of Care for Outpatients with Heart Failure." *Annals of Internal Medicine* 146 (4): 270–7.

Blumenthal, D. 1993. "Total Quality Management and Physicians' Clinical Decisions." *JAMA* 269 (21): 2775–8.

———. 2009. "Stimulating the Adoption of Health Information Technology." *New England Journal of Medicine* 360 (15): 1477–9.

———. 2010. "Launching HITECH." *New England Journal of Medicine* 362 (5): 382–5.

Borzecki, A. M., A. T. Wong, E. C. Hickey, A. S. Ash, and D. R. Berlowitz. 2004. "Can We Use Automated Data to Assess Quality of Hypertension Care?" *American Journal of Managed Care* 10 (7 Pt 2): 473–9.

Care Transformation Chronic Care Sustainability Initiative. 2017. "Care Transformation Collaborative Rhode Island" [accessed on March 20, 2017]. Available at https://www.ctc-ri.org/

Centers for Medicare and Medicaid Services. 2016. "Quality Initiatives - General Information" [accessed on March 20, 2017]. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/index.html

Centers for Medicare and Medicaid Services. 2017. "Certified EHR Technology" [accessed on August 23, 2017]. Available at https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Certification.html

Centers for Medicare and Medicaid Services. "Value Based Payment Modifier" [accessed on March 20, 2017]. Available at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/ValueBased Payment
Modifier.html

Danford, C. P., A. M. Navar-Boggan, J. Stafford, C. McCarver, E. D. Peterson, and T. Y. Wang. 2013. "The Feasibility and Accuracy of Evaluating Lipid Management Performance Metrics Using an Electronic Health Record." *American Heart Journal* 166 (4): 701–8.

DeShazo, J. P., and M. A. Hoffman. 2015. "A Comparison of a Multistate Inpatient EHR Database to the HCUP Nationwide Inpatient Sample." *BMC Health Services Research* 15 (1): 1.

DeVoe, J. E., R. Gold, P. McIntire, J. Puro, S. Chauvie, and C. A. Gallia. 2011. "Electronic Health Records vs Medicaid Claims: Completeness of Diabetes Preventive Care Data in Community Health Centers." *Annals of Family Medicine* 9 (4): 351–8.

Heintzman, J., S. R. Bailey, M. J. Hoopes, T. Le, R. Gold, J. P. O'Malley, S. Cowburn, M. Marino, A. Krist, and J. E. DeVoe. 2014. "Agreement of Medicaid Claims and Electronic Health Records for Assessing Preventive Care Quality among Adults." *Journal of the American Medical Informatics Association* 21 (4): 720–4.

Jollis, J. G., M. Ancukiewicz, E. R. DeLong, D. B. Pryor, L. H. Muhlbaier, and D. B. Mark. 1993. "Discordance of Databases Designed for Claims Payment Versus Clinical Information Systems. Implications for Outcomes Research." *Annals of Internal Medicine* 119 (8): 844–50.

Lohr, K. N. 1990. "Use of Insurance Claims Data in Measuring Quality of Care." *International Journal of Technology Assessment in Health Care* 6 (2): 263–71.

Maclean, J. R., D. M. Fick, W. K. Hoffman, C. T. King, E. R. Lough, and J. L. Waller. 2002. "Comparison of Two Systems for Clinical Practice Profiling in Diabetic Care: Medical Records Versus Claims and Administrative Data." *American Journal of Managed Care* 8 (2): 175–9.

Mainous 3rd, A. G., and J. Talbert. 1998. "Assessing Quality of Care Via HEDIS 3.0. Is There a Better Way?" *Archives of Family Medicine* 7 (5): 410–3.

McCoy, R. G., S. M. Tulledge-Scheitel, J. M. Naessens, A. E. Glasgow, R. J. Stroebel, S. J. Crane, K. S. Bunkers, and N. D. Shah. 2016. "The Method for Performance Measurement Matters: Diabetes Care Quality as Measured by Administrative Claims and Institutional Registry." *Health Services Research* 51: 2206–20.

Office of the National Coordinator for Health Information Technology. 2014. "Health IT Legislation and Regulations" [accessed on March 20, 2017]. Available at https://www.healthit.gov/policy-researchers-implementers/health-it-legislation-and-regulations

Parsons, A., C. McCullough, J. Wang, and S. Shih. 2012. "Validity of Electronic Health Record-Derived Quality Measurement for Performance Monitoring." *Journal of the American Medical Informatics Association* 19 (4): 604–9.

Porter, M. E., S. Larsson, and T. H. Lee. 2016. "Standardizing Patient Outcomes Measurement." *New England Journal of Medicine* 374 (6): 504–6.

Rhode Island Quality Institute. 2017. "Rhode Island Quality Institute: About" [accessed on March 20, 2017]. Available at http://www.riqi.org/matriarch/default.html

Tang, P. C., M. Ralston, M. F. Arrigotti, L. Qureshi, and J. Graham. 2007. "Comparison of Methodologies for Calculating Quality Measures Based on Administrative Data Versus Clinical Data from an Electronic Health Record System: Implications for Performance Measures." *Journal of the American Medical Informatics Association* 14 (1): 10–5.

Urech, T. H., L. D. Woodard, S. S. Virani, R. A. Dudley, M. Z. Lutschg, and L. A. Petersen. 2015. "Calculations of Financial Incentives for Providers in a Pay-for-Performance Program: Manual Review Versus Data from Structured Fields in Electronic Health Records." *Medical Care* 53 (10): 901–7.

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.