Published in final edited form as:
Am J Prev Med. 2016 November ; 51(5): 752–761. doi:10.1016/j.amepre.2016.07.004.

Measuring Preventive Care Delivery:

Comparing Rates Across Three Data Sources

Steffani R. Bailey, PhD[1], **John D. Heintzman, MD, MPH**[1], **Miguel Marino, PhD**[1,2], **Megan J. Hoopes, MPH**[3], **Brigit A. Hatch, MD**[1,3], **Rachel Gold, PhD, MPH**[3,4], **Stuart C. Cowburn, MPH**[3], **Christine A. Nelson, PhD, RN**[3], **Heather E. Angier, MPH**[1], and **Jennifer E. DeVoe, MD, DPhil**[1,3]

[1]Department of Family Medicine, Oregon Health and Science University, Portland, Oregon

[2]Department of Public Health and Preventive Medicine, Division of Biostatistics, Oregon Health and Science University, Portland, Oregon

[3]OCHIN, Inc., Portland, Oregon

[4]Kaiser Permanente Northwest Center for Health Research, Portland, Oregon

Abstract

Introduction—Preventive care delivery is an important quality outcome, and electronic data reports are increasingly being used to track these services. It is highly informative when electronic data sources are compared to information manually extracted from medical charts to assess validity and completeness.

Methods—This cross-sectional study used a random sample of Medicaid-insured patients seen at 43 community health centers in 2011 to calculate standard measures of correspondence between manual chart review and two automated sources (electronic health records [EHRs] and Medicaid claims), comparing documentation of orders for and receipt of ten preventive services (n=150 patients/service). Data were analyzed in 2015.

Results—Using manual chart review as the gold standard, automated EHR extraction showed near-perfect to perfect agreement (κ=0.96–1.0) for services received within the primary care setting (e.g., BMI, blood pressure). Receipt of breast and colorectal cancer screenings, services commonly referred out, showed moderate (κ=0.42) to substantial (κ=0.62) agreement, respectively. Automated EHR extraction showed near-perfect agreement (κ=0.83–0.97) for

Address correspondence to: Steffani R. Bailey, PhD, Oregon Health and Science University, Department of Family Medicine, 3181 SW Sam Jackson Park Rd., Portland OR 97329. bailstef@ohsu.edu.

All authors approved the final version of the manuscript and revised the manuscript for intellectual content. SRB, JDH, and JED made substantial contributions to the study design, data interpretation, and initial drafting of the manuscript. MM, MJH, and SCC contributed to the data analysis and interpretation. BAH, RG, CAN, and HEA assisted with study design and data interpretation.

The article contents were presented at the American Medical Informatics Association 2015 Annual Symposium in San Francisco, CA, November 14–18.

No financial disclosures were reported by the authors of this paper.

documentation of ordered services. Medicaid claims showed near-perfect agreement (κ=0.87) for hyperlipidemia and diabetes screening, and substantial agreement (κ=0.67–0.80) for receipt of breast, cervical, and colorectal cancer screenings, and influenza vaccination. Claims showed moderate agreement (κ=0.59) for chlamydia screening receipt. Medicaid claims did not capture ordered or unbilled services.

Conclusions—Findings suggest that automated EHR and claims data provide valid sources for measuring receipt of most preventive services; however, ordered and unbilled services were primarily captured via EHR data and completed referrals were more often documented in claims data.

Introduction

Preventive care delivery is an important quality outcome in value-based care models.[1–3] With the 2015 passage of the Medicare Access and Children's Health Insurance Program (CHIP) Reauthorization Act, increasing emphasis will be placed on preventive service delivery and population health outcomes. Data reports that extract information from electronic health records (EHRs) and health insurance claims are increasingly used to track such measures of preventive care quality.[1,2,4–7] Federal programs, notably the Centers for Medicare and Medicaid Services' "Meaningful Use" (MU) of EHRs Incentive Program, accelerated these efforts. It is meaningful for electronic data sources to be compared to manually extracted medical chart data to assess validity and completeness. The validity of data used to track receipt of preventive services has important implications. It is also useful to assess preventive care offered to the patient, such as orders for screening tests, particularly in settings such as community health centers (CHCs) where patients routinely face barriers to follow-through on referrals.

Before electronic data sources were used, delivery of preventive services was measured by manually extracting information from patients' medical charts. In research, manual chart review is often still considered the "gold standard,"[8–12] though this method is time consuming and expensive and thus can only be applied to small numbers of patients. As electronic systems matured, administrative health insurance claims were increasingly used to capture information on larger populations. Claims data, however, do not capture unbilled services or services provided to uninsured patients. EHRs, now commonly used, have the potential to supplement (or supplant) claims data as a more complete electronic source of information. Although many systems have switched to using EHRs, especially in light of MU requirements, little is known about how different data sources compare in the quality of their data on preventive care. Previous comparisons of two data sources (e.g., EHR versus manual chart review,[10,13–19] or EHR versus administrative claims[11,12,20–28]) typically focused on a limited number of measures and yielded mixed results. Few studies have used specifications from Stage 1 MU of EHRs to assess provision of a broad range of recommended preventive services,[10,24] and none have compared both EHR extraction and claims data to manual chart review.

In this cross-sectional study, EHR data and Medicaid administrative claims data were extracted via automated processes and compared to manual chart review (data abstracted from individual patient charts) among continuously Medicaid-insured adults served by a network of CHCs. Patient-level agreement was assessed on documentation of receipt of ten preventive services that were part of Stage 1 MU[2] or recommended in national guidelines.[29,30] Documentation of whether preventive services were ordered in the two data sets containing this information (EHR, manual chart review) was also compared. It was hypothesized that EHR data extracted using automated processes would have higher agreement with manual chart review than data from Medicaid claims. If automated EHR data have high agreement with manual chart review, automated EHR data may be less time intensive and more cost effective than manual review, and would allow for real-time assessment of receipt of preventive care, regardless of insurance coverage.

Methods

Study Population

Oregon Medicaid enrollment data were used to identify patients aged 19–64 years, who were continuously insured by Medicaid throughout 2011, and had one or more billing claim from an Oregon OCHIN CHC in 2011. Medicaid identification numbers were used to match patients across data sets. Among matched patients, those with one or more primary care encounter in one or more of the 43 Oregon CHCs that implemented the OCHIN EHR before January 1, 2010 (to ensure use of the EHR for ≥1 year prior to data collection; N=18,471 patients) were identified. Patients were excluded who had evidence of any insurance coverage other than Medicaid (n=3,870), were pregnant (n=1,494), or died (n=6) in 2011. The resulting data set included 13,101 patients who appeared in both the OCHIN EHR and Oregon Medicaid claims data sets.

Sample size calculations[31] were performed based on an expected kappa statistic of 0.65, a prevalence of 30% receipt in preventive services (as a conservative estimate) and a minimum difference of ±0.10 between the kappa statistic and its lower (or upper) 95% confidence bound. Based on these calculations, 150 patients eligible for each preventive service (described below) were randomly sampled; patients could be in more than one denominator (i.e., the same patient might be randomly chosen for inclusion in the blood pressure subsample and the hyperlipidemia screening subsample).

This study was approved by the IRB of Oregon Health and Science University.

Data Sample

The EHR data were obtained from OCHIN (formerly the Oregon Community Health Information Network; now “OCHIN,” as other states joined), a non-profit community health information network providing a single, linked Epic© EHR to CHCs.[32–34] “Ordered” and “received” were calculated separately for services that would likely be ordered and completed at different times (e.g., cholesterol screening) or for services commonly referred out (e.g., mammogram, colorectal cancer screening).

For each measure, standardized manual data collection algorithms that utilized discrete and free text fields and scanned documents were created, to include data inaccessible via electronic extraction. Reviewers entered data into a secure data management system formatted using Research Electronic Data Capture (REDCap) software.[35] Individual patient charts were reviewed by two OCHIN staff members by outcome. Both were trained by a physician researcher who uses Epic© charting in clinical practice. Prior to review, a random subset of charts (n=20 for each outcome) was evaluated by each OCHIN abstractor and the physician researcher to confirm inter-rater reliability. The two reviewers had perfect agreement for most outcomes (screenings for breast cancer, hyperlipidemia, diabetes; BMI, and blood pressure assessments) and >75% agreement on all remaining measures.

Using structured query language coding, data were extracted on preventive services from the source EHR database. This automated EHR extraction included data captured in discrete fields only (e.g., diagnoses, procedure orders, medication orders) based on standardized code sets. For the EHR automated queries, codes were based on Stage 1 MU measures, developed/implemented in 2011, when the ordered and received services were documented.[2] These included ICD-9-CM diagnosis codes, Current Procedural Terminology and Healthcare Common Procedure Coding System codes, Logical Observation Identifiers Names and Codes codes, and medication codes. Relevant codes and groupings specific to the OCHIN EHR for internal reporting,[24] and discrete fields that capture data entered into the EHR via “check boxes,” typically used to record receipt of patient-reported services with an associated date (e.g., influenza vaccination received at another facility), were also included.

Enrollment and claims data were obtained for all patients insured by Oregon’s Medicaid program for 2011. This data set was obtained >18 months after the end of the measurement year to account for lag time in claims processing. Codes used to capture service provision in these data were based on the Healthcare Effectiveness Data and Information Set specifications, which are tailored to claims-based reporting[3] and included standard diagnosis, procedure, and revenue codes. The physician measures did not include specifications for influenza vaccination, blood pressure assessment, or hyperlipidemia screening, so the code sets used for assessing these measures in claims were the same as those used for the automated EHR data extraction.

Measures

Within each of the data sources, documentation of orders for/receipt of ten recommended adult preventive care services[29,30] during 2011 were assessed—screening for cervical, breast, and colorectal (i.e., colonoscopy, fecal occult blood test, or flexible sigmoidoscopy) cancers; screening for chlamydia, hyperlipidemia, and diabetes (hemoglobin A1c); blood pressure, BMI, and smoking status assessments; and receipt of influenza vaccination. The intent of this analysis was to compare the data sets in their documentation of preventive service delivery for a given patient, not to identify who should have received that service based on time of last receipt. Thus, patient eligibility for a given preventive service was based on recommended age, sex, and history of previous procedures (Table 1 footnotes), but not whether the patient was due for a service.

Statistical Analysis

Study sample demographics using EHR data were described, and documentation of each preventive service in each of the three data sources was assessed. First, the percentage of eligible patients (per sex, age, and relevant medical history) with documented services in the automated EHR extraction, manual chart review, and in Medicaid claims were tabulated; proportions were compared between manual chart review and each automated data source using McNemar's test of agreement. Then, common statistical measures of correspondence were calculated at the patient level between the manual chart review and (1) automated EHR extraction, and (2) Medicaid claims: *sensitivity* (proportion of patients where the comparator data set denoted "ordered" or "received" when manual chart review did the same), *specificity* (proportion of comparator patients correctly classified as "not received" when manual chart review denoted not having received a service), *positive predictive value* (probability of a patient having received the service when the data source denoted "received"), *negative predictive value* (probability of a patient not having received the service when the data source denoted "not received"), *agreement* (total proportion of patients in which the compared data sets denote the same status), and *kappa* statistic (similar to agreement, but removes agreement that would be expected purely by chance), with exact 95% CIs where appropriate.[36] Kappa values of 0.00–0.20 were considered slight agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1.00 as near-perfect agreement.[37] Analyses were performed in 2015 using SAS, version 9.4.

Results

Demographic characteristics of the 150 randomly sampled patients for each outcome are shown in Table 1 (n=1,113 unique individuals, as patients could be in more than one denominator). The EHR automated extraction documented a significantly lower number of patients receiving breast and colorectal cancer screenings compared with manual chart review (Table 2). Medicaid claims data recorded significantly lower numbers of patients receiving BMI assessments and influenza vaccinations, and higher numbers of receipt of hyperlipidemia and chlamydia screenings, as compared with data from manual chart review.

At the patient level (i.e., service documentation for a given patient), EHR data obtained via automated extraction agreed with manual chart review perfectly for BMI and blood pressure assessment (Table 3). Near-perfect agreement (>92%, κ>0.8) was observed for hyperlipidemia and diabetes screening, influenza vaccination, chlamydia screening, and cervical cancer screening, and substantial agreement for smoking assessment and receipt of colorectal cancer screening (>90%, κ>0.60). The lowest agreement was for breast cancer screening documented as "received" (80%, κ=0.42), although documentation of "ordered" mammography showed near-perfect agreement (98%, κ=0.95). The automatically extracted EHR data correctly identified patients for whom services were received (positive predictive value ≥0.94 for all; sensitivity ≥0.8 for all measures except received breast and colorectal cancer screening [sensitivity of 0.36 and 0.50, respectively]). Automatically extracted EHR data also had high specificity (≥0.92) and negative predictive value (≥0.93) for all measures except smoking status assessment (0.67 specificity, 0.71 negative predictive value).

Near-perfect agreement (≥94%, κ>0.81) was observed between Medicaid claims and manual chart review at the patient level for receipt of hyperlipidemia and diabetes screening, and substantial agreement for receipt of breast cancer screening (91.3%, κ=0.80), cervical cancer screening (86.7%, κ=0.67), colorectal cancer screening (93.3%, κ=0.79), and influenza vaccination (88.0%, κ=0.74). Chlamydia screening had moderate agreement (79.3%, κ=0.59). Few patients had BMI, blood pressure, or smoking assessment identifiable in the claims data, thus these measures had very low agreement. Excluding services that usually do not generate a claim and documentation, Medicaid claims performed well on both identifying patients who received services (sensitivity ≥0.75 for all) and those who did not (specificity ≥0.71 for all). In most cases, these measures of performance were somewhat lower for the Medicaid claims comparison than for the automatically extracted EHR data.

Discussion

This is the first study to examine agreement between manual chart review and two electronic sources in assessing preventive care provision in primary care. When measuring rates of receipt of referred preventive services (as a percentage), claims data sources appeared more complete than the manual chart review and the automated EHR data (Table 2). However, the three data sources were more similar in agreement of receipt at the patient level (Table 3). Overall, there was high agreement between the EHR data and manual chart review. Claims data had moderate to near-perfect agreement with manual chart review for many services; however, clinical quality measures that do not often generate a separate claim (e.g., assessing smoking status, blood pressure, BMI) were not well documented in the claims data. The ability to electronically extract these measures from the EHR in a way that shows near-perfect to perfect agreement with manual review supports the use of electronic methods that extract EHR data to measure these and other unbilled services much more efficiently than manual review.

Agreement was higher between the manual chart review and EHR data than between manual chart review and claims data for receipt of services usually performed in the clinic (e.g., cervical cancer screening, hyperlipidemia screening). Consistent with previous findings, services that are often referred out of the primary care setting (e.g., breast and colorectal cancer screenings) were more commonly documented in Medicaid claims.[24] In the EHR, documentation of completed mammography and colonoscopy, for example, are often returned as scanned documents, which cannot be accessed using automated queries and would only be found upon manual review. Previous studies also noted this limitation of automated EHR data extraction, especially when information is not contained in structured fields.[10,14,24] To maximize the use of EHRs for reporting receipt of services, improved processes that expand capture of services rendered outside of the primary care clinic are warranted. These improvements might include changes in the clinical workflow, data entry tools that force information into discrete data fields, or data extraction procedures such as natural language processing.[38,39] The lack of systemwide integration of EHRs and immature health information exchanges in the U.S. also limit the ability to assess services obtained outside of a specified network. Based on findings from a previous analysis, the integration of EHR and claims data could provide the most comprehensive assessment of healthcare quality. However, this “hybrid” method would be limited to patients with a single payer source; given the necessity to match patients in both data sources and the lag time from service receipt to insurance claims submission and approval, is likely not an efficient long-term solution that would produce “real-time” data for supporting clinical decisions.[24]

This study highlights a limitation of using insurance claims data as the basis for assessing provision of preventive services: claims data capture completed services only, thus missing services that were ordered but not completed. Measuring performance based solely on completed screenings could lead to inadvertent bias against providers who order appropriate screening tests, but whose patients face multiple barriers to follow through on such referrals. Consequently, these providers might not meet requirements for incentive payments, despite ordering recommended preventive care. Knowledge of whether a test was ordered but not received also provides a critical opportunity for outreach to patients to encourage follow-through. In this study, when the metric of interest was whether the provider attempted to provide the service (i.e., ordered the screening), automated EHR data extraction had high agreement with manual chart review. Future studies should examine the differences in care quality if the metric is ordered versus received care, particularly among uninsured populations.

Although automated EHR data extraction queries are more efficient than manual chart reviews for large populations, identifying the code sets and developing the data extraction process can take considerable time and resources; such technologies are still being developed and not currently available as “out-of-the-box” options from most EHR vendors. Potential financial barriers to developing automated EHR extraction processes could be offset by the incentives offered by Centers for Medicare and Medicaid Services’ MU program, and standardized metrics could result in specific code sets used to assess each outcome. In addition, these costs may be worth the investment as an increasing percentage of payers transition to value-based reimbursement models and healthcare systems will need the most accurate data sources possible to track and improve their care delivery.

Limitations

Only patients who were continuously insured by Medicaid and received services at Oregon CHCs were included in this study, which may limit generalizability. It could not be determined from Medicaid claims the location where services were received; thus, some services documented in the claims data could have been received somewhere other than an OCHIN CHC. The intent of the current analysis was to conduct a patient-level comparison of services documented in 1 year in each data set; whether the patient was due for the services was not assessed. Therefore, the reported rates should not be compared to national care quality rates. Finally, the data extraction procedure was conducted using data from one EHR system that allowed for query of the backend database, and was performed by an analyst with experience in conducting these queries; future studies are needed to determine whether other systems are capable of these queries and whether these results can be replicated.

Conclusions

The current findings suggest that data automatically extracted from an EHR is valid for evaluating and documenting preventive care, particularly ordered services and services received within the primary care setting, whereas Medicaid claims are better at capturing services that are referred out of the primary care setting. Although both the automated EHR and Medicaid claims data sources agreed with the manual chart review for many preventive services, Medicaid claims do not include documentation of ordered screenings, and often do not capture receipt of unbilled services, many of which are included as MU measures or other national recommendations for preventive care quality. As primary care clinics must monitor both billed and unbilled services across patient populations with variable insurance statuses, it is essential to have valid technologies to electronically extract these data. There is also a need for continued focus on best practices for EHR-based electronic information exchanges between the primary care and specialty providers for referred preventive services.

Acknowledgments

The authors gratefully acknowledge the OCHIN network and the OCHIN PBRN for making this research possible. We would like to acknowledge Allison Casciato, MD, for her assistance with the REDCap database setup; Tessa Carter for her assistance in the manual chart review; and Lorie Jacob, ScM, for statistical assistance.

This work was supported by grant R01HL107647 from the National Heart, Lung, and Blood Institute (registered as an observational study in clinicaltrials.gov, Identifier NCT02355132), grant UL1TR000128 from the National Center for Advancing Translational Sciences, grant K23DA037453 from the National Institute on Drug Abuse, and grant UL1RR024140 from the National Center for Research Resources. The funding agencies had no role in study design; collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication. The research presented in this paper is that of the authors and does not reflect the official policy of NIH.

References

1. DHHS, Centers for Medicare and Medicaid Services. Medicare and Medicaid programs; electronic health record incentive program. Final rule. Fed Regist. 2010; 75(144):44313–44588. [PubMed: 20677415]
2. Center for Medicaid and Medicare Services. 2011-2012 Eligible Professional Clinical Quality Measures (CQMs). 2012. https://ushik.ahrq.gov/mdr/portals/mu. Accessed January 26, 2016
3. National Committee for Quality Assurance (NCQA). HEDIS 2013: Technical Specifications for Physician Measurement. Washington, DC: 2013.
4. Weiner JP, Fowles JB, Chan KS. New paradigms for measuring clinical performance using electronic health records. Int J Qual Health Care. 2012; 24(3):200–205. http://dx.doi.org/10.1093/intqhc/mzs011. [PubMed: 22490301]
5. Kern LM, Barron Y, Dhopeshwarkar RV, Edwards A, Kaushal R. Electronic health records and ambulatory quality of care. J Gen Intern Med. 2013; 28(4):496–503. http://dx.doi.org/10.1007/s11606-012-2237-8. [PubMed: 23054927]
6. Lobach DF, Detmer DE. Research challenges for electronic health records. Am J Prev Med. 2007; 32(5 Suppl):S104–111. http://dx.doi.org/10.1016/j.amepre.2007.01.018. [PubMed: 17466814]
7. Baer HJ, Cho I, Walmer RA, Bain PA, Bates DW. Using electronic health records to address overweight and obesity: a systematic review. Am J Prev Med. 2013; 45(4):494–500. http://dx.doi.org/10.1016/j.amepre.2013.05.015. [PubMed: 24050426]
8. Tinoco A, Evans RS, Staes CJ, Lloyd JF, Rothschild JM, Haug PJ. Comparison of computerized surveillance and manual chart review for adverse events. J Am Med Inform Assoc. 2011; 18(4): 491–497. http://dx.doi.org/10.1136/amiajnl-2011-000187. [PubMed: 21672911]
9. Vinson DR, Morley JE, Huang J, et al. The Accuracy of an Electronic Pulmonary Embolism Severity Index Auto-Populated from the Electronic Health Record: Setting the stage for computerized clinical decision support. Appl Clin Inform. 2015; 6(2):318–333. http://dx.doi.org/10.4338/ACI-2014-12-RA-0116. [PubMed: 26171078]
10. Kern LM, Kaushal R. Accuracy of electronically reported "meaningful use" clinical quality measures. Ann Intern Med. 2013; 159(1):73. http://dx.doi.org/10.7326/0003-4819.159-1-201307020-00017.

11. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010; 67(5):503–527. http://dx.doi.org/10.1177/1077558709359007. [PubMed: 20150441]

12. Tang PC, Ralston M, Arrigotti MF, Qureshi L, Graham J. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. J Am Med Inform Assoc. 2007; 14(1):10–15. http://dx.doi.org/10.1197/jamia.M2198. [PubMed: 17068349]

13. Kerr EA, Smith DM, Hogan MM, et al. Comparing clinical automated, medical record, and hybrid data sources for diabetes quality measures. Jt Comm J Qual Improv. 2002; 28(10):555–565. [PubMed: 12369158]

14. Baker DW, Persell SD, Thompson JA, et al. Automated review of electronic health records to assess quality of care for outpatients with heart failure. Ann Intern Med. 2007; 146(4):270–277. http://dx.doi.org/10.7326/0003-4819-146-4-200702200-00006. [PubMed: 17310051]

15. Kmetik KS, O'Toole MF, Bossley H, et al. Exceptions to outpatient quality measures for coronary artery disease in electronic health records. Ann Intern Med. 2011; 154(4):227–234. http://dx.doi.org/10.7326/0003-4819-154-4-201102150-00003. [PubMed: 21320938]

16. Persell SD, Wright JM, Thompson JA, Kmetik KS, Baker DW. Assessing the validity of national quality measures for coronary artery disease using an electronic health record. Arch Intern Med. 2006; 166(20):2272–2277. http://dx.doi.org/10.1001/archinte.166.20.2272. [PubMed: 17101947]

17. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. J Am Med Inform Assoc. 2012; 19(4):604–609. http://dx.doi.org/10.1136/amiajnl-2011-000557. [PubMed: 22249967]

18. Agnew-Blais JC, Coblyn JS, Katz JN, Anderson RJ, Mehta J, Solomon DH. Measuring quality of care for rheumatic diseases using an electronic medical record. Ann Rheum Dis. 2009; 68(5):680–684. http://dx.doi.org/10.1136/ard.2008.089318. [PubMed: 18511547]

19. Benin AL, Vitkauskas G, Thornquist E, et al. Validity of using an electronic medical record for assessing quality of care in an outpatient setting. Med Care. 2005; 43(7):691–698. http://dx.doi.org/10.1097/01.mlr.0000167185.26058.8e. [PubMed: 15970784]

20. Devoe JE, Gold R, McIntire P, Puro J, Chauvie S, Gallia CA. Electronic health records vs Medicaid claims: completeness of diabetes preventive care data in community health centers. Ann Fam Med. 2011; 9(4):351–358. http://dx.doi.org/10.1370/afm.1279. [PubMed: 21747107]

21. MacLean CH, Louie R, Shekelle PG, et al. Comparison of administrative data and medical records to measure the quality of medical care provided to vulnerable older patients. Med Care. 2006; 44(2):141–148. http://dx.doi.org/10.1097/01.mlr.0000196960.12860.de. [PubMed: 16434913]

22. Linder JA, Kaleba EO, Kmetik KS. Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. Med Care. 2009; 47(2):208–216. http://dx.doi.org/10.1097/MLR.0b013e318189375f. [PubMed: 19169122]

23. Rosenman MB, Tao G, Szucs KA, Wang J, Ambuehl R, Mahon BE. Prenatal syphilis screening rates measured using medicaid claims and electronic medical records. Sex Transm Dis. 2008; 35(4):387–392. http://dx.doi.org/10.1097/OLQ.0b013e31815fa5bb. [PubMed: 18362860]

24. Heintzman J, Bailey SR, Hoopes MJ, et al. Agreement of Medicaid claims and electronic health records for assessing preventive care quality among adults. J Am Med Inform Assoc. 2014; 21(4):720–724. http://dx.doi.org/10.1136/amiajnl-2013-002333. [PubMed: 24508767]

25. Kottke TE, Baechler CJ, Parker ED. Accuracy of heart disease prevalence estimated from claims data compared with an electronic health record. Prev Chronic Dis. 2012; 9:E141. http://dx.doi.org/10.5888/pcd9.120009. [PubMed: 22916996]

26. Solberg LI, Engebretson KI, Sperl-Hillen JM, Hroscikoski MC, O'Connor PJ. Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and depression. Am J Med Qual. 2006; 21(4):238–245. http://dx.doi.org/10.1177/1062860606288243. [PubMed: 16849780]

27. Angier H, Gold R, Gallia C, et al. Variation in outcomes of quality measurement by data source. Pediatrics. 2014; 133(6):e1676–1682. [PubMed: 24864178]

28. Gold R, Angier H, Mangione-Smith R, et al. Feasibility of Evaluating the CHIPRA Care Quality Measures in Electronic Health Record Data. Pediatrics. 2012; 130(1):139–149. http://dx.doi.org/10.1542/peds.2011-3705. [PubMed: 22711724]
29. Centers for Disease Control and Prevention. Summary Recommendations: Prevention and Control of Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices—(ACIP)—United States, 2013-14. 2013. www.cdc.gov/flu/professionals/acip/2013-summary-recommendations.htm. Accessed January 26, 2016
30. U.S. Preventive Services Task Force. USPSTF A and B Recommendations. 2014. www.uspreventiveservicestaskforce.org/Page/Name/uspstf-a-and-b-recommendations/. Accessed January 26, 2016
31. Cantor A. Sample-size calculations for Cohen's kappa. Psychol Methods. 1996; 1(2):150–153. http://dx.doi.org/10.1037/1082-989X.1.2.150.
32. OCHIN. OCHIN 2012 Annual Report. 2013. https://ochin.org/depository/OCHIN%202012%20Annual%20Report.pdf. Accessed January 26, 2016
33. Devoe JE, Sears A. The OCHIN community information network: bringing together community health centers, information technology, and data to support a patient-centered medical village. J Am Board Fam Med. 2013; 26(3):271–278. http://dx.doi.org/10.3122/jabfm.2013.03.120234. [PubMed: 23657695]
34. Devoe JE, Gold R, Spofford M, et al. Developing a network of community health centers with a common electronic health record: description of the Safety Net West Practice-based Research Network (SNW-PBRN). J Am Board Fam Med. 2011; 24(5):597–604. http://dx.doi.org/10.3122/jabfm.2011.05.110052. [PubMed: 21900444]
35. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009; 42(2):377–381. http://dx.doi.org/10.1016/j.jbi.2008.08.010. [PubMed: 18929686]
36. Morissette JT, Khorram S. Exact binomial confidence interval for proportions. Photogramm Eng Remote Sens. 1998; 64(4):281–282.
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1):159–174. http://dx.doi.org/10.2307/2529310. [PubMed: 843571]
38. Gao H, Aiello Bowles EJ, Carrell D, Buist DS. Using natural language processing to extract mammographic findings. J Biomed Inform. 2015; 54:77–84. http://dx.doi.org/10.1016/j.jbi.2015.01.010. [PubMed: 25661260]
39. Denny JC, Choma NN, Peterson JF, et al. Natural language processing improves identification of colorectal cancer testing in the electronic medical record. Med Decis Making. 2012; 32(1):188–197. http://dx.doi.org/10.1177/0272989X11400418. [PubMed: 21393557]

Table 1

Demographic Distribution of Study Sample by Measure

Measure (N=150 patients per measure)	Age mean (SD)	Male N (%)	White race N (%)	Hispanic ethnicity N (%)	Primary language English N (%)	FPL ≤138% N (%)
BMI assessment	41.4 (11.9)	64 (42.7)	125 (83.3)	19 (12.7)	128 (85.3)	141 (94.0)
Blood pressure assessment	41.0 (12.0)	49 (32.7)	129 (86.0)	23 (15.3)	127 (84.7)	140 (93.3)
Smoking status assessment	40.7 (12.5)	54 (36.0)	129 (86.0)	16 (10.7)	134 (89.3)	143 (95.3)
Hyperlipidemia screening[a]	39.8 (12.9)	49 (32.7)	124 (82.7)	19 (12.7)	133 (88.7)	143 (95.3)
Diabetes screening[b]	52.5 (5.2)	63 (42.0)	126 (84.0)	19 (12.7)	130 (86.7)	145 (96.7)
Influenza vaccination[c]	56.0 (3.9)	62 (41.3)	111 (74.0)	19 (12.7)	121 (80.7)	142 (94.7)
Chlamydia screening[d]	21.5 (1.4)	--	98 (65.3)	30 (20.0)	128 (85.3)	143 (95.3)
Cervical cancer screening[e]	38.6 (12.8)	--	134 (89.3)	24 (16.0)	128 (85.3)	143 (95.3)
Breast cancer screening[f]	50.4 (6.5)	--	132 (88.0)	15 (10.0)	133 (88.7)	146 (97.3)
Colorectal cancer screening[g]	55.3 (3.8)	50 (33.3)	127 (84.7)	17 (11.3)	129 (86.0)	145 (96.7)

[a]Males and females ages 20-64

[b]Males and females ages 45-64

[c]Males and females ages 50-64

[d]Sexually active females ages 19-24

[e]Females ages 19-64 with no history of hysterectomy

[f]Females ages 40-64 with no history of bilateral mastectomy

[g]Males and females ages 50-64 with no history of colorectal cancer or total colectomy FPL, federal poverty level

Table 2

Percent of Eligible Patients[a] Recorded as Positive For Service (Ordered or Received), by Data Source[b]

Measure (N=150 patients per measure)	Manual chart review (%)	EHR automated extraction (%)	Medicaid claims (%)
BMI assessment received	88.0	88.0	**0.7** **
Blood pressure assessment received	99.3	99.3	0
Smoking status assessment received	90.0	90.7	0
Hyperlipidemia screening[c]			
Ordered	34.7	32.7	--
Received	32.0	32.7	**36.7** *
Diabetes screening[d]			
Ordered	39.3	41.3	--
Received	37.3	38.0	39.3
Influenza vaccination[e], received	40.0	39.3	**32.0** *
Chlamydia screening[f]			
Ordered	39.3	40.7	--
Received	40.0	40.7	**54.0** **
Cervical cancer screening[g]			
Ordered	28.7	24.7	--
Received	27.3	26.0	28.7
Breast cancer screening[h]			
Ordered	28.7	26.7	--
Received	30.0	**11.3** **	33.3
Colorectal cancer screening[i]			
Ordered	26.7	23.3	--
Received	18.7	**9.33** **	20.0

Notes: -- Not applicable; claims data measure only services received

[*] Boldface indicates statistical significance $p<0.05$, McNemar test of agreement exact test, vs manual chart review

[**] Boldface indicates statistical significance $p<0.001$, McNemar test of agreement exact test, vs manual chart review

[a] Eligibility was based recommended age, sex, and medical history, not whether patient was due for a service

[b] Percentages do not account for whether service provision documentation is for the same patient across data sources

[c] Males and females ages 20-64; cholesterol screening includes LDL, HDL, total cholesterol, and triglycerides

[d] Males and females ages 45-64

[e] Males and females ages 50-64

[f] Sexually active females ages 19-24

[g]Females ages 19-64 with no history of hysterectomy

[h]Females ages 40-64 with no history of bilateral mastectomy

[i]Males and females ages 50-64 with no history of colorectal cancer or total colectomy; colorectal cancer screening includes fecal occult blood test, flexible sigmoidoscopy, and colonoscopy EHR, electronic health record

Table 3

Measures of Correspondence For Comparisons of Manual Chart Review vs. Automated Extraction of EHR Data and Medicaid Claims Data

Measure (N=150 patients per measure)	Agreement (%)	Kappa (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
BMI assessment						
MCR vs. EHR, received	100	1.0	1.0	1.0	1.0	1.0
MCR vs. claims, received	12.7	0.002 (−0.002, 0.006)	0.008 (0.0002, 0.04)	1.0	1.0	0.12 (0.07, 0.18)
Blood pressure assessment						
MCR vs. EHR, received	100	1.0	1.0	1.0	1.0	1.0
MCR vs. claims, received	0.7	--	--	1.0	--	0.007 (0.0002, 0.04)
Smoking status assessment						
MCR vs. EHR, received	94.0	0.66 (0.45, 0.86)	0.97 (0.93, 0.99)	0.67 (0.38, 0.88)	0.96 (0.92, 0.99)	0.71 (0.42, 0.92)
MCR vs. claims, received	10.0	--	--	1.0	--	0.10 (0.06, 0.16)
Hyperlipidemia screening[a]						
MCR vs. EHR, ordered	94.0	0.87 (0.78, 0.95)	0.88 (0.77, 0.96)	0.97 (0.91, 0.99)	0.94 (0.83, 0.99)	0.94 (0.88, 0.98)
MCR vs. EHR, received	99.3	0.98 (0.96, 1.0)	1.0	0.99 (0.95, 1.0)	0.98 (0.89, 1.0)	1.0
MCR vs. claims, received	94.0	0.87 (0.78, 0.95)	0.98 (0.89, 1.0)	0.92 (0.85, 0.97)	0.85 (0.73, 0.94)	0.99 (0.94, 1.0)
Diabetes screening[b]						
MCR vs. EHR, ordered	92.7	0.85 (0.76, 0.93)	0.93 (0.84, 0.98)	0.92 (0.85, 0.97)	0.89 (0.78, 0.95)	0.95 (0.89, 0.99)
MCR vs. EHR, received	98.0	0.96 (0.91, 1.0)	0.98 (0.90, 1.0)	0.98 (0.93, 1.0)	0.96 (0.88, 1.0)	0.99 (0.94, 1.0)
MCR vs. claims, received	94.0	0.87 (0.79, 0.95)	0.95 (0.85, 0.99)	0.94 (0.87, 0.98)	0.89 (0.79, 0.96)	0.97 (0.91, 0.99)
Influenza vaccination[c]						
MCR vs. EHR, received	98.0	0.96 (0.91, 1.0)	0.97 (0.88, 1.0)	0.99 (0.94, 1.0)	0.98 (0.91, 1.0)	0.98 (0.92, 1.0)
MCR vs. claims, received	88.0	0.74 (0.63, 0.85)	0.75 (0.62, 0.85)	0.97 (0.91, 0.99)	0.94 (0.83, 0.99)	0.85 (0.77, 0.92)
Chlamydia screening[d]						
MCR vs. EHR, ordered	98.7	0.97 (0.93, 1.0)	1.0	0.98 (0.92, 1.0)	0.97 (0.89, 1.0)	1.0
MCR vs. EHR, received	98.0	0.96 (0.91, 1.0)	0.98 (0.91, 1.0)	0.98 (0.92, 1.0)	0.97 (0.89, 1.0)	0.99 (0.94, 1.0)
MCR vs. claims, received	79.3	0.59 (0.47, 0.72)	0.92 (0.82, 0.97)	0.71 (0.61, 0.80)	0.68 (0.57, 0.78)	0.93 (0.84, 0.98)
Cervical cancer screening[e]						
MCR vs. EHR, ordered	93.3	0.83 (0.73, 0.93)	0.81 (0.67, 0.92)	0.98 (0.93, 1.0)	0.95 (0.82, 0.99)	0.93 (0.87, 0.97)
MCR vs. EHR, received	97.3	0.93 (0.87, 1.0)	0.93 (0.80, 0.98)	0.99 (0.95, 1.0)	0.97 (0.86, 1.0)	0.97 (0.92, 1.0)
MCR vs. claims, received	86.7	0.67 (0.54, 0.80)	0.78 (0.62, 0.89)	0.90 (0.83, 0.95)	0.74 (0.59, 0.86)	0.92 (0.85, 0.96)
Breast cancer screening[f]						

Measure (N=150 patients per measure)	Agreement (%)	Kappa (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
MCR vs. EHR, ordered	98.0	0.95 (0.89, 1.0)	0.93 (0.81, 0.99)	1.0	1.0	0.97 (0.92, 0.99)
MCR vs. EHR, received	80.0	0.42 (0.27, 0.57)	0.36 (0.22, 0.51)	0.99 (0.95, 1.0)	0.94 (0.71, 1.0)	0.78 (0.70, 0.85)
MCR vs. claims, received	91.3	0.80 (0.70, 0.90)	0.91 (0.79, 0.98)	0.91 (0.84, 0.96)	0.82 (0.69, 0.91)	0.96 (0.90, 0.99)
Colorectal cancer screening[g]						
MCR vs. EHR, ordered	96.7	0.91 (0.84, 0.99)	0.88 (0.73, 0.96)	1.0	1.0	0.96 (0.90, 0.99)
MCR vs. EHR, received	90.7	0.62 (0.44, 0.79)	0.50 (0.31, 0.69)	1.0	1.0	0.90 (0.83, 0.94)
MCR vs. claims, received	93.3	0.79 (0.66, 0.91)	0.85 (0.67, 0.96)	0.95 (0.90, 0.98)	0.80 (0.61, 0.92)	0.97 (0.92, 0.99)

Notes: 95% CIs were obtained using exact binomial CIs for proportions

-- Statistic cannot be computed due to zero cell count in claims data Eligibility was based recommended age, sex, and medical history, not whether patient was due for a service

EHR, electronic health record; MCR, manual chart review; PPV, positive predictive value; NPV, negative predictive value

[a]Males and females ages 20-64; cholesterol screening includes LDL, HDL, total cholesterol, and triglycerides

[b]Males and females ages 45-64

[c]Males and females ages 50-64

[d]Sexually active females ages 19-24

[e]Females ages 19-64 with no history of hysterectomy

[f]Females ages 40-64 with no history of bilateral mastectomy

[g]Males and females ages 50-64 with no history of colorectal cancer or total colectomy; colorectal cancer screening includes fecal occult blood test, flexible sigmoidoscopy, and colonoscopy