



Medicaid coverage accuracy in electronic health records

Miguel Marino^{a,b,*}, Heather Angier^a, Steele Valenzuela^a, Megan Hoopes^c, Marie Killerby^a, Brenna Blackburn^a, Nathalie Huguet^a, John Heintzman^a, Brigit Hatch^{a,c}, Jean P. O'Malley^{a,c}, Jennifer E. DeVoe^{a,c}

^a Department of Family Medicine, Oregon Health & Science University, Portland, OR, USA

^b School of Public Health, Oregon Health & Science University, Portland, OR, USA

^c OCHIN, Portland, OR, USA

ARTICLE INFO

Keywords:

Electronic health records
Medicaid
Health policy
Health insurance

ABSTRACT

Health insurance coverage facilitates access to preventive screenings and other essential health care services, and is linked to improved health outcomes; therefore, it is critical to understand how well coverage information is documented in the electronic health record (EHR) and which characteristics are associated with accurate documentation. Our objective was to evaluate the validity of EHR data for monitoring longitudinal Medicaid coverage and assess variation by patient demographics, visit types, and clinic characteristics. We conducted a retrospective, observational study comparing Medicaid status agreement between Oregon community health center EHR data linked at the patient-level to Medicaid enrollment data (gold standard). We included adult patients with a Medicaid identification number and ≥ 1 clinic visit between 1/1/2013–12/31/2014 [> 1 million visits ($n = 135,514$ patients)]. We estimated statistical correspondence between EHR and Medicaid data at each visit (visit-level) and for different insurance cohorts over time (patient-level). Data were collected in 2016 and analyzed 2017–2018. We observed excellent agreement between EHR and Medicaid data for health insurance information: kappa (> 0.80), sensitivity (> 0.80), and specificity (> 0.85). Several characteristics were associated with agreement; at the visit-level, agreement was lower for patients who preferred a non-English language and for visits missing income information. At the patient-level, agreement was lower for black patients and higher for older patients seen in primary care community health centers. Community health center EHR data are a valid source of Medicaid coverage information. Agreement varied with several characteristics, something researchers and clinic staff should consider when using health insurance information from EHR data.

1. Introduction

Health insurance coverage facilitates access to preventive screenings and other essential health care services, and is linked to improved health outcomes (Kasper et al., 2000; Asplin et al., 2005; O'Malley et al., 2016; Hatch et al., 2017). Thus, understanding how accurate documentation of this information is in the electronic health record (EHR), and identifying characteristics associated with accurate documentation, are critical. Research has in recent years increasingly used EHR data due to its widespread uptake (Jones and Furukawa, 2014; Heisey-Grove and Patel, 2015) and rich, longitudinal, patient-level information. Validation of EHR data is therefore needed to confidently use it for health services research, as the data are collected for clinical and billing, not research, purposes. Previous studies have validated EHR data to assess preventive care, quality measure, risk stratification, and performance

metrics, and for conducting health insurance surveillance (Heintzman et al., 2014a; Bailey et al., 2016; Hirsch and Scheck McAlearney, 2014; Heintzman et al., 2014b; Hatch et al., 2013, 2016; Gold et al., 2012; Kharrazi et al., 2017). For example, one study found that EHR data of influenza vaccination and cholesterol and cervical cancer screening rates had excellent agreement with Medicaid claims (Heintzman et al., 2014a). Excellent insurance agreement between EHR and Medicaid data (i.e., the proportion of visits for which both EHR and Medicaid data denoted the same coverage) was also reported for pediatric health insurance information (Heintzman et al., 2015); however, this analysis did not examine agreement for adults, nor identify the multi-level factors associated with agreement. Medicaid continuity may differ between adults and children due to disparate eligibility and re-enrollment requirements (Henry J. Kaiser Family Foundation, 2014) and dissimilar patterns of health care receipt. In addition, disease capture between

* Corresponding author at: Department of Family Medicine, Division of Biostatistics, School of Public Health, Oregon Health & Science University, 3181 SW Sam Jackson Park Rd, Portland, OR 97239, USA.

E-mail address: marinom@ohsu.edu (M. Marino).

<https://doi.org/10.1016/j.pmedr.2018.07.009>

Received 19 February 2018; Received in revised form 19 July 2018; Accepted 21 July 2018

Available online 27 July 2018

2211-3355/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

EHR and claims data is impacted by number of visits and continuous health insurance (Kottke et al., 2012; Devoe et al., 2011a); yet, little is known about factors associated with accurate capture of insurance coverage status in the EHR.

Many EHR systems cannot be linked to insurance enrollment records and only contain records of insurance status on the date of service. Therefore, it is critical to understand how well the health insurance information collected from the EHR corresponds to true enrollment status, and which characteristics are associated with accurate documentation. This study: (i) compares Medicaid insurance agreement between EHR data from a network of Oregon community health centers (CHCs) and Medicaid enrollment data (the ‘gold standard’) and (ii) describes characteristics associated with agreement. We hypothesized that health insurance information from the EHR would have excellent agreement with Medicaid enrollment data.

2. Methods

2.1. Data sources

This retrospective, observational study used EHR data from Oregon CHCs in the OCHIN, Inc. health information network, and Medicaid enrollment data (referred to as EHR data and Medicaid data, respectively). OCHIN hosts a single instance of the Epic® EHR (Epic Systems Inc., Verona, WI) (DeVoe et al., 2011b) which contains comprehensive information on patient demographics, visit and clinic characteristics, billing and primary payer recorded at each visit. Primary payer information is collected by the front desk at each visit; however, it is limited to the date a patient receives services, and patients may be uncertain about their enrollment status. Medicaid enrollment data came from the Medicaid Management Information System of the Oregon Health Authority, and contain the actual date ranges of Medicaid coverage for individual patients. Patient-level linkages were created between the EHR dataset and the Medicaid dataset using each individual patient's Oregon Medicaid Identification (ID) number, a unique identifier appearing in both data sources. CHCs are a good setting for validation of Medicaid coverage in the EHR since the majority of their patients have Medicaid (National Association of Community Health Centers, 2017). Furthermore, CHCs commonly assist their patients with health insurance enrollment and retention (<https://www.hrsa.gov/about/news/press-releases/2015-11-13-outreach-enrollment.html>, 2015), a unique function which supports the need for high-quality insurance coverage information in the EHR.

2.2. Study population and time period

We included patients aged 19–64 years with either an Oregon Medicaid ID or Medicaid insurance recorded in the OCHIN EHR with ≥ 1 billed health care visit (excluding dental) from 184 Oregon CHCs in the OCHIN network linked to state Medicaid records during the study time period (1/1/2013–12/31/2014). Patients were linked using their Oregon Medicaid ID, which appeared in both data sources. As this study spans the start of the Affordable Care Act (ACA) Medicaid expansion and Oregon expanded Medicaid up to 138% of the federal poverty level (FPL), we split the study time period for analyses into pre-ACA [1/1/2013–12/31/2013 (N = 89,305 patients)] and post-ACA [1/1/2014–12/31/2014 (N = 109,883 patients)]. We assessed the pre- and post-ACA time periods separately because we hypothesized that the dramatic increase in coverage due to the ACA (Oregon Medicaid enrollment increased 43.7% from 2013 to March 2014) (<http://kff.org/medicaid/issue-brief/how-is-the-aca-impacting-medicaid-enrollment/>, 2014) could impact agreement rates.

2.3. Visit- and patient-level outcomes

We considered two types of analyses commonly performed in health

services research: visit- and patient-level. For visit-level analyses, we addressed whether EHR and Medicaid data denoted the same Medicaid insurance status at each visit. Patients could qualify for the study in one or both of the pre-/post-ACA time periods. Visit-level coverage was evaluated by two methods: 1) primary payer recorded at the visit (Medicaid or not Medicaid) from EHR data; and 2) enrollment dates from Medicaid data (visits were considered Medicaid if they fell between coverage start and end dates and not Medicaid if the visit was outside the coverage period).

For patient-level analyses, we addressed whether insurance cohorts (based on longitudinal insurance coverage patterns) agreed across the two data sources. We restricted inclusion in the insurance cohorts (described below) to patients who had ≥ 1 visit pre- and ≥ 1 visit post-ACA (N = 63,674 patients). Cohorts were identified separately using EHR data and Medicaid data utilizing definitions from prior studies (O'Malley et al., 2016; Gold et al., 2014):

- 1) Continuously Medicaid: Medicaid recorded at all visits;
- 2) Continuously not Medicaid: Medicaid not recorded at any visit (patients could have Medicare, private, VA/Military, worker's compensation, or no coverage);
- 3) Gained Medicaid: Medicaid not recorded for all visits in 2013 and Medicaid recorded at all visits in 2014; and
- 4) Discontinuously Medicaid: Any combination of visit coverage that did not follow the definitions above.

The primary outcome was agreement between EHR and Medicaid data in assigning patients to one of the four cohorts.

2.4. Covariates

The following EHR-derived covariates were hypothesized as potential influences on agreement and considered in analyses: patient demographics (sex, age, race, ethnicity, language, household income represented as % FPL, urban/rural, number of common chronic conditions, and number of encounters) (Kottke et al., 2012), visit and provider types, and clinic-specific (department type and customers of OCHIN's billing service) characteristics. We included the following common chronic conditions assessed from diagnostic codes in the EHR: hypertension, diabetes, coronary artery disease, lipid disorder, and asthma/chronic obstructive pulmonary disorder.

2.5. Statistical analyses

For all analyses, we compared EHR data to the ‘gold standard’ of Medicaid data. For visit-level analyses, we calculated common statistical correspondence measures: agreement, prevalence-adjusted bias-adjusted kappa (PABAK), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Agreement was defined as the proportion of visits in which both EHR and Medicaid data denoted the same coverage. PABAK agreement, which overcomes the limitation of kappa being highly dependent on prevalence, adjusts for the prevalence of the outcome and removes agreement expected due to chance (Byrt et al., 1993). The range for low agreement is 0–0.40, moderate agreement is 0.41–0.60, substantial agreement is 0.61–0.79 and excellent agreement is 0.80–1.00 (Viera and Garrett, 2005). As visits were clustered within patients, confidence intervals for all correspondence measures were estimated using nonparametric cluster bootstrapping with 5000 repeats (Field and Welsh, 2007).

To investigate characteristics associated with agreement in visit-level analyses, we used a two-stage logistic regression model that estimates odds ratios of agreement controlling for agreement due to chance (Lipsitz et al., 2003). The first stage consisted of separate standard logistic regressions for EHR and Medicaid data. This stage estimated an offset utilized in the second stage to control for agreement due to chance. In the second stage, a single logistic regression model was performed.

Table 1
Characteristics of patients and visits.

Characteristics	Study population for visit-level analysis		Study population for patient-level analysis
	Patient/visit characteristics for visits occurring in 2013	Patient/visit characteristics for visits occurring in 2014	Patient/visit characteristics for patients with visits in 2013 and 2014
Patient demographics	N = 89,305	N = 109,883	N = 63,674
Sex, N (%)			
Male	29,562 (33.1)	40,239 (36.6)	20,726 (32.5)
Female	59,739 (66.9)	69,638 (63.4)	43,037 (67.5)
Age, mean (SD)	38.8 (12.8)	38.7 (12.9)	39.7 (12.6)
Race, N (%)			
White	74,382 (83.3)	91,927 (83.7)	53,757 (84.3)
Black	5109 (5.7)	5601 (5.1)	3500 (5.5)
Other	5730 (6.4)	7035 (6.4)	4050 (6.4)
Unknown	4084 (4.6)	5320 (4.8)	2457 (3.9)
Ethnicity, N (%)			
Hispanic	17,202 (19.3)	19,646 (17.9)	12,805 (20.1)
Non-Hispanic	69,137 (77.4)	86,377 (78.6)	49,396 (77.5)
Unknown	2966 (3.3)	3860 (3.5)	1563 (2.5)
Preferred language, N (%)			
English	71,656 (80.2)	89,312 (81.3)	50,758 (79.6)
Spanish	12,026 (13.5)	13,277 (12.0)	9182 (14.4)
Other	4850 (5.4)	5746 (5.2)	3488 (5.5)
Unknown	773 (0.9)	1598 (1.5)	336 (0.5)
Federal poverty level, N (%)			
> 200%	4169 (4.7)	5113 (4.7)	3037 (4.8)
139–200%	5663 (6.3)	7352 (6.7)	4255 (6.6)
< 139%	73,651 (82.5)	88,938 (80.9)	52,780 (82.8)
Missing	5822 (6.5)	8480 (7.7)	3722 (5.8)
Most frequented health center location, N (%)			
Urban	80,858 (90.5)	99,317 (90.4)	57,230 (89.8)
Rural	8130 (9.1)	10,129 (9.2)	6305 (9.9)
Missing	317 (0.4)	437 (0.4)	229 (0.4)
Number of visits, N (%)			
< 6	26,512 (29.7)	45,530 (41.4)	10,602 (16.6)
6–20	45,776 (51.3)	48,036 (43.7)	37,303 (58.5)
> 20	17,017 (19.1)	16,317 (14.8)	15,859 (24.9)
Number of chronic conditions, N (%)			
0	52,493 (58.8)	68,559 (62.4)	33,998 (53.3)
1	19,694 (22.1)	23,164 (21.1)	15,412 (24.2)
> 1	17,118 (19.2)	18,160 (16.5)	14,354 (22.5)

Characteristics	Study population for visit-level analysis		Study population for patient-level analysis
	Patient/visit characteristics for visits occurring in 2013	Patient/visit characteristics for visits occurring in 2014	Patient/visit characteristics for patients with visits in 2013 and 2014
Visit types, N	N = 468,699	N = 538,658	N = 751,224
Number of visits covered by Medicaid, as reported by Medicaid data, N (%)	293,208 (62.6)	476,306 (88.4)	556,055 (74.0)
Number of visits that identified Medicaid, as reported by EHR, N (%)	245,919 (52.5)	409,340 (76.0)	465,041 (61.9)
Visit coverage type, as reported by EHR, N (%)			
Medicaid	245,919 (52.5)	409,340 (76.0)	465,041 (61.9)
Medicare	48,140 (10.3)	46,375 (8.6)	82,146 (10.9)
Private	28,376 (6.1)	23,500 (4.4)	40,688 (5.4)
Self-pay (no coverage)	143,638 (30.6)	56,575 (10.5)	159,208 (21.2)
Visit type, N (%)			
Medical office visit	320,194 (68.3)	368,158 (68.3)	515,467 (68.6)
Obstetrics	14,254 (3.0)	13,033 (2.4)	17,972 (2.4)
Mental, behavioral, or case management	38,047 (8.1)	35,218 (6.5)	56,712 (7.5)
Lab, imaging, or immunization-only	30,924 (6.6)	39,819 (7.4)	51,669 (6.9)
All other types	65,280 (13.9)	82,430 (15.3)	109,404 (14.6)

(continued on next page)

Table 1 (continued)

Characteristics	Study population for visit-level analysis		Study population for patient-level analysis
	Patient/visit characteristics for visits occurring in 2013	Patient/visit characteristics for visits occurring in 2014	Patient/visit characteristics for patients with visits in 2013 and 2014
Visit types, N	N = 468,699	N = 538,658	N = 751,224
Provider type, N (%)			
Medical Doctor	172,963 (36.9)	192,811 (35.8)	279,694 (37.2)
Mid-level provider	158,351 (33.8)	188,741 (35.0)	249,865 (33.3)
Mental health/behavioral health provider	44,016 (9.4)	41,166 (7.6)	67,217 (8.9)
Specialist	11,380 (2.4)	8575 (1.6)	15,899 (2.1)
Support staff/other	72,105 (15.4)	91,702 (17.0)	121,186 (16.1)

Note: OCHIN health information network Epic® EHR data are referred to as EHR data and Oregon Medicaid enrollment data are referred to as Medicaid data. Demographics were extracted from EHR patient enrollment tables. The number of chronic conditions were determined from diagnosis codes in the patients' medical record and ranged from 0 to 5 based on the following conditions: hypertension, diabetes, coronary artery disease, lipid disorder, and asthma/chronic obstructive pulmonary disorder. Visit type was grouped using primary level of service CPT codes (e.g., 99,201–99,205) and visit type as coded in the EHR. Provider types were extracted from EHR provider data. Data were collected in 2016 and analyzed 2017–2018.

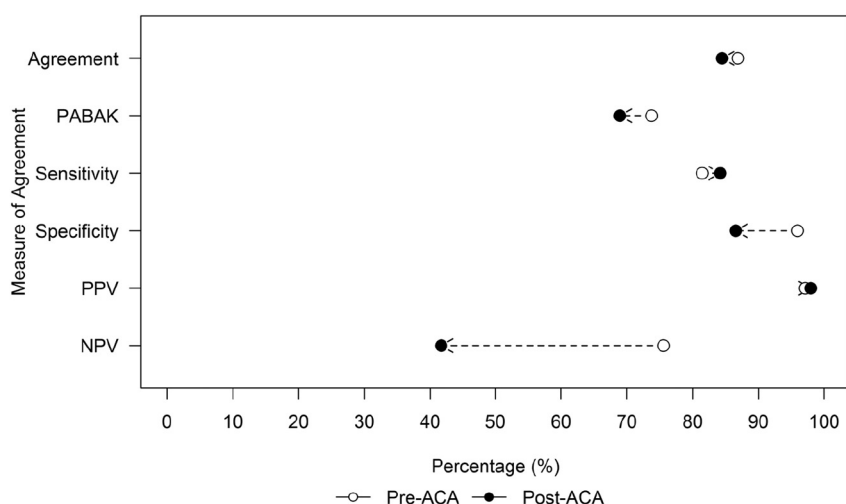


Fig. 1. Measures of Medicaid coverage agreement between EHR and Medicaid data, stratified by pre- and post-Affordable Care Act (ACA) periods.

Note: OCHIN health information network Epic® electronic health record (EHR) data are referred to as EHR data and Oregon Medicaid enrollment data are referred to as Medicaid data. Agreement is defined as total proportion of encounters in which EHR data denoted the same coverage status as the 'gold standard' (i.e., Medicaid data). PABAK adjusts kappa for differences in prevalence of the conditions and for bias between data sources. Sensitivity is the probability that EHR denoted coverage when the assumed gold standard also denoted coverage. Specificity is the probability that EHR correctly classified 'no Medicaid coverage' when the assumed gold standard also denotes no Medicaid coverage. PPV is the likelihood of a visit being covered by Medicaid when the gold standard denoted Medicaid coverage. NPV is the likelihood of an encounter not being covered by Medicaid when the gold standard denoted no Medicaid coverage. Data were collected in 2016 and analyzed 2017–2018.

For patient-level analyses, we produced a 4-by-4 cross-tabulation of insurance cohorts by data source and estimated agreement and kappa statistics. To investigate characteristics associated with agreement, we considered an extension of the two-stage logistic regression model described above for the insurance cohorts. The first stage consisted of separate multinomial logistic regression for each data source to obtain marginal probabilities of being assigned to a given cohort to estimate the offset needed to adjust for chance agreement. The second stage implemented the offsets into a single logistic regression model. For all analyses, we included the covariates listed above at both stages of the regression modeling and cluster bootstrapping with 5000 repeats to account for visits nested within patients and patients within clinics. Additional modeling details are in the Appendix.

Data were collected in 2016 and analyses performed during 2017–2018. Analyses were performed using R.v.3.2.5 (R Development Team). This study was approved by the Institutional Review Board at Oregon Health & Science University.

3. Results

The total number of patients increased from 89,305 to 109,883 from pre- to post-ACA (Table 1). Overall, patients in the study sample were predominantly female, white, non-Hispanic, in households earning ≤138% of the FPL, and received care in CHCs in urban areas. The majority of patients in the sample did not have documentation of common chronic medical conditions. Of the 468,699 visits pre-ACA, 62.6% had Medicaid coverage according to Medicaid data. The number

of visits increased to 538,658 post-ACA, and Medicaid data indicated that 88.4% of those visits were covered by Medicaid.

3.1. Visit-level analyses

Measures of correspondence between EHR and Medicaid data at the visit-level had excellent agreement (> 0.80), PABAK (> 0.65), sensitivity (> 0.80), specificity (> 0.85), and PPV (> 0.95) both pre- and post-ACA (Fig. 1). Only specificity and NPV were different when comparing correspondence statistics pre- and post-ACA (both decreased post-ACA). We produced a histogram of clinic agreement for both periods demonstrating excellent agreement > 80% in most clinics (Appendix Fig. A.1).

At the visit-level (Table 2), the odds of agreement were higher for males pre-ACA [odds ratio (OR) = 1.14, 95% confidence interval (CI) = 1.07–1.23] compared to females. The odds of agreement were lower for visits in rural compared to urban CHCs in pre-ACA (OR = 0.86, 95% CI = 0.78–0.94); however, this significant difference was no longer observed post-ACA (OR = 1.03, 95% CI = 0.96–1.10). All patients with non-English language preference saw lower odds of agreement both pre- and post-ACA than those with English language preference; odds were less pronounced post-ACA. Visits with missing income information had lower odds of agreement both pre-and post-ACA, while higher patient FPL was associated with greater odds of agreement post-ACA only. Miscellaneous and obstetrics visit types had lower odds of agreement compared to lab/imaging/immunization-only visits.

Table 2
Odds ratio of agreement between EHR and Medicaid data for *patient-, visit-, and clinic-level* characteristics associated with *visit-level* Medicaid coverage.

Characteristics	Pre-ACA visits (1/1/2013–12/31/2013)	Post-ACA visits (1/1/2014–12/31/2014)
	OR (95% CI)	OR (95% CI)
Patient demographics		
Sex		
Female	Ref	Ref
Male	1.14 (1.07–1.22)	1.03 (0.98–1.08)
Age		
19–29	Ref	Ref
30–39	1.13 (1.05–1.22)	1.12 (1.07–1.18)
40–49	1.35 (1.25–1.49)	1.14 (1.06–1.22)
50–64	1.22 (1.11–1.33)	1.21 (1.14–1.30)
Race		
White	Ref	Ref
Black	0.92 (0.80–1.06)	0.99 (0.90–1.11)
Other	1.00 (0.88–1.16)	0.98 (0.87–1.09)
Unknown	1.04 (0.92–1.18)	0.98 (0.90–1.06)
Ethnicity		
Non-Hispanic	Ref	Ref
Hispanic	0.91 (0.81–1.02)	1.02 (0.94–1.11)
Unknown	0.87 (0.71–1.11)	0.88 (0.77–1.05)
Language		
English	Ref	Ref
Spanish	0.45 (0.40–0.50)	0.79 (0.71–0.86)
Other	0.78 (0.66–0.92)	0.81 (0.72–0.91)
FPL		
< 139%	Ref	Ref
139–200%	1.23 (1.10–1.40)	1.44 (1.32–1.57)
> 200%	1.09 (0.96–1.24)	1.58 (1.43–1.74)
Missing	0.44 (0.40–0.48)	0.90 (0.84–0.98)
Urban		
Urban	Ref	Ref
Rural	0.86 (0.78–0.93)	1.03 (0.96–1.10)
Number of chronic conditions		
0	Ref	Ref
1	1.11 (1.03–1.19)	1.07 (1.01–1.14)
> 1	1.01 (0.93–1.11)	1.16 (1.09–1.24)
No. of encounters		
< 6	Ref	Ref
6–20	0.97 (0.92–1.02)	1.00 (0.96–1.04)
> 20	0.69 (0.65–0.74)	0.92 (0.87–0.97)
Visit types		
Visit type		
Medical office visit	Ref	Ref
Mental, behavioral, or case management	1.24 (1.11–1.38)	1.03 (0.93–1.13)
Misc.	0.77 (0.72–0.83)	0.82 (0.78–0.87)
Lab, imaging, or immunization-only	0.99 (0.91–1.09)	1.13 (1.05–1.21)
Obstetrics	0.39 (0.36–0.43)	0.90 (0.81–1.01)
Provider type		
Medical doctor	Ref	Ref
Mid-level provider	1.21 (1.16–1.27)	1.03 (1.00–1.07)
Mental health/behavior health provider	0.94 (0.85–1.05)	0.96 (0.87–1.05)
Specialist	0.49 (0.43–0.55)	0.77 (0.70–0.85)
Support staff/other	1.15 (1.06–1.26)	0.86 (0.80–0.92)
Clinic characteristics		
Department type		
Primary care	Ref	Ref
Medical specialty	0.69 (0.56–0.86)	0.79 (0.70–0.89)
Mental health	0.42 (0.36–0.48)	0.67 (0.57–0.79)
Public health	0.97 (0.82–1.20)	1.04 (0.92–1.18)
OCHIN billing service customer		
No	Ref	Ref
Yes	1.20 (1.13–1.28)	0.92 (0.89–0.97)

Note: OCHIN health information network Epic® EHR data are referred to as EHR data and Oregon Medicaid enrollment data are referred to as Medicaid data. Bolded numbers denote statistical significance. We used the following abbreviations: ACA = Affordable Care Act; OR = odds ratio; CI = confidence

interval; Ref = reference category. OCHIN billing services is a business line that OCHIN offers to CHC members where OCHIN conduct all the billing needs on behalf of the CHC instead of doing it themselves. We hypothesized that this would equate to more uniform and expert billing & collection practices. Data were collected in 2016 and analyzed 2017–2018.

3.2. Patient-level analyses

Overall, patient-level analyses showed EHRs correctly classified 78.8% of the Continuously Medicaid cohort and 82.3% the Gained Medicaid cohort (Appendix Fig. A.2). The EHR was less able to accurately capture patients in the Continuously not Medicaid or Discontinuous Medicaid cohorts (43.7% and 41.5%, respectively). Nearly 30% of the Discontinuous Medicaid group (per Medicaid data) was reported as Continuously Medicaid in the EHR data, and over 50% of patients reported as Continuously not Medicaid in Medicaid data were reported as having Medicaid at some point in the EHR.

At the patient-level (Fig. 2), the odds of agreement between EHR and Medicaid data were lower for patients who were black, preferred Spanish language, preferred other (non-English, non-Spanish) language, missing FPL, and patients with > 5 encounters compared to their reference groups. We observed higher odds of agreement for patients in older age groups, with ≥ 138% of the FPL, with at least one chronic condition, seen in a primary care CHC, and whose clinic used billing services hosted by OCHIN.

4. Discussion

EHR data are increasingly used for health services research; their growth and expansion have been suggested as a source for public health surveillance, preventive services outreach, and social determinants of health tracking and action planning (Gottlieb et al., 2015; Klompas et al., 2012; Remington and Wadland, 2015). Therefore, it is important to ensure that the data are valid. In this study, we found excellent agreement between EHR and Medicaid data which confirms previous studies demonstrating the accuracy of EHR data for research (O'Malley et al., 2016; Hirsch and Scheck McAlearney, 2014; Hatch et al., 2013; Heintzman et al., 2015; Klompas et al., 2012; Hoopes et al., 2016), and expands confidence in EHR data for studies addressing adults and their Medicaid enrollment.

Our findings also support the conduct of research using EHR data in the CHC setting. As over 26 million patients, many of whom belong to vulnerable populations, are seen at CHCs (National Association of Community Health Centers, 2017), the use of EHR data for research may also be an important resource for population health planning and health equity initiatives moving forward. Correct insurance information in CHCs is also important as many CHCs offer insurance enrollment services to help their patients gain or maintain insurance (<https://www.hrsa.gov/about/news/press-releases/2015-11-13-outreach-enrollment.html>, 2015; Hall et al., 2017; DeVoe et al., 2013, 2014a; Harding et al., 2017). We found that EHR and Medicaid data agreement varied with patient demographics, visit types, and clinic characteristics; agreement also varied by these characteristics depending on the time period. For example, patients living in rural areas pre-ACA had significantly lower odds of agreement than those living in urban areas; however, the difference was not significant post-ACA. It could be that the ACA had a larger impact on Medicaid coverage for those living in rural compared to urban areas (Soni et al., 2017) and this greater influx of enrollees in rural areas changed the odds of agreement post-ACA when compared to patients living in urban areas. Thus, despite excellent agreement, researchers and clinic staff should be aware of possible differences in coverage agreement based on patient demographics, visit types, and clinic characteristics, and clinicians should understand that the accuracy of their patients' recorded data may also be affected by these factors. Because insurance status facilitates financial accessibility to preventive services (Heintzman et al., 2014b),

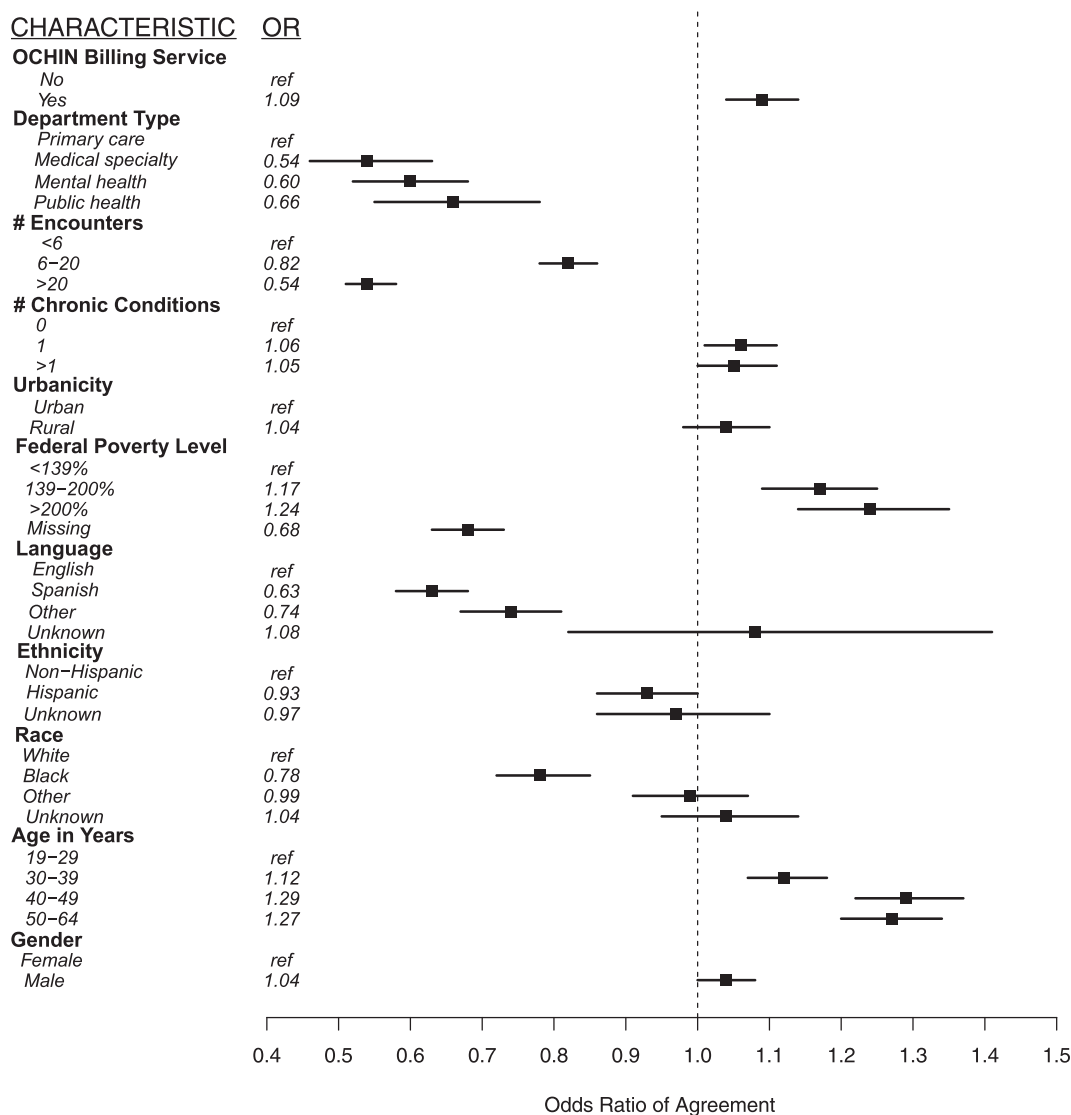


Fig. 2. Odds ratios for patient and clinic factors associated with agreement on Medicaid coverage at the patient-level, between EHR and Medicaid data. Note: OCHIN health information network Epic© EHR data are referred to as EHR data and Oregon Medicaid enrollment data are referred to as Medicaid data. The number of chronic conditions ranged from 0 to 5 based on the following conditions: hypertension, diabetes, coronary artery disease, lipid disorder, and asthma/chronic obstructive pulmonary disorder. Data were collected in 2016 and analyzed 2017–2018.

the accuracy of the data may also have profound impacts on a patient's ability to access essential preventive care.

Our finding that more visits were associated with decreased agreement does not follow prior disease-based analyses (Kottke et al., 2012). Yet, health insurance can change at each visit whereas disease states often remain stable, and differences in agreement for patients with many visits may be related to a higher chance for error. The inaccuracies with patients at lower FPL may be a result of assumptions that all patients earning ≤ 138% FPL were insured by Medicaid, when some of these patients may not have been eligible for Medicaid or enrolled in Medicaid for other reasons.

While the Continuous Medicaid and Gained Medicaid groups had excellent agreement, the agreement was much lower in the Continuous not Medicaid and Discontinuous Medicaid groups. There are several potential reasons for this. Nearly 30% of the Discontinuous Medicaid group (per Medicaid data) was reported as Continuously Medicaid in the EHR data, likely because patients did not have health care visits while uninsured. Therefore, the EHR data would not have any record of the patient's insurance gap. Over 50% of patients reported as Continuously not Medicaid in Medicaid data were reported as having

Medicaid at some point in the EHR. This difference in agreement could potentially stem from workflows that fail to accurately verify coverage for certain patients; insurance status and eligibility may not be documented uniformly across all patient groups, especially non-English speakers. Additional research is needed to understand whether these differences are due to workflow, insurer factors, or how coverage information is transmitted to clinics. Clinics could use these findings as the basis to explore other strategies for improving the accuracy of patient health insurance information in the EHR, such as building direct linkages with Medicaid and other payers to keep information accurate and even build reminder systems for patients whose coverage has lapsed or is nearing expiration (Hall et al., 2017; DeVoe et al., 2014b; Gold et al., 2015). Because of the established link between preventive care and health insurance status, such efforts could substantially bolster the receipt of preventive care of economically vulnerable populations.

This study has limitations. Although our study included a large sample of patients from 184 CHCs, it was limited to one state and one networked EHR system. It is probable that other EHRs record health insurance information differently. State Medicaid programs also differ in eligibility and enrollment procedures. The methods we described

here should be applied to other states and EHR systems with different populations and insurance types for wider applicability. Another limitation is that the selection criteria for patients may skew the population towards those with health insurance, as they have been observed to seek health care services more than those without coverage. However, CHCs see patients regardless of insurance status, so this effect may be smaller in the CHC setting than other health care settings.

5. Conclusion

Community health center EHR data are a valid source of Medicaid coverage information. Despite excellent agreement, researchers and clinics should take into consideration differences in coverage agreement based on patient demographics, visit types, and clinic characteristics, and consider measures to further improve data accuracy to inform efforts to increase preventive screening rates and the delivery of other essential health care services.

Acknowledgments

The authors acknowledge the significant contributions to this study that were provided by collaborating investigators in the NEXT-D2 (Natural Experiments in Translation for Diabetes) Study Two and the OCHIN clinics. This study was supported by the Agency for Healthcare Research and Quality (grant #R01HS024270), the National Heart, Lung and Blood Institute (grant #R01HL136575) and the National Cancer Institute (grant #s R01CA204267 and R01CA181452). This publication was also made possible by Cooperative Agreement Number U18DP006116 jointly funded by the U.S. Centers for Disease Control and Prevention and the National Institute of Diabetes and Digestive and Kidney Disease, and Patient-Centered Outcomes Research Institute (PCORI). The authors also acknowledge the participation of our partnering health systems. The views expressed do not necessarily reflect the official policies of the DHHS or at PCORI nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government. OHSU IRB# IRB00011858. We thank Amanda Delzer Hill for her editing work that greatly improved this manuscript.

Previous presentations

The content of this paper was presented at the Western North American Region of The International Biometric Society Annual Meeting (June 2017, Santa Fe, NM) and the North American Primary Care Research Group Annual Meeting (November 2016, Colorado Springs, CO).

Conflict of interest

No financial disclosures were reported by the authors of this paper. The funding agency had no role in study design; collection, analysis, and interpretation of data; writing or submission for publication.

Appendix A. Supplementary materials

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pmedr.2018.07.009>.

References

Asplin, B.R., Rhodes, K.V., Levy, H., et al., 2005. Insurance status and access to urgent ambulatory care follow-up appointments. *JAMA* 294, 1248–1254.
 Bailey, S.R., Heintzman, J.D., Marino, M., et al., 2016. Measuring preventive care delivery: comparing rates across three data sources. *Am. J. Prev. Med.* 51, 752–761.
 Byrt, T., Bishop, J., Carlin, J.B., 1993. Bias, prevalence and kappa. *J. Clin. Epidemiol.* 46, 423–429.

DeVoe, J.E., 2013. Being uninsured is bad for your health: can medical homes play a role in treating the uninsured ailment? *Ann. Fam. Med.* 11, 473–476.
 DeVoe, J.E., Gold, R., McIntire, P., Puro, J., Chauvie, S., Gallia, C.A., 2011a. Electronic health records vs Medicaid claims: completeness of diabetes preventive care data in community health centers. *Ann. Fam. Med.* 9, 351–358.
 DeVoe, J.E., Gold, R., Spofford, M., et al., 2011b. 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.* 24, 597–604.
 DeVoe, J., Angier, H., Likumahwa, S., et al., 2014a. Use of qualitative methods and user-centered design to develop customized health information technology tools within federally qualified health centers to keep children insured. *J. Ambul. Care Manage.* 37, 148–154.
 DeVoe, J.E., Angier, H., Burdick, T., Gold, R., 2014b. Health information technology: an untapped resource to help keep patients insured. *Ann. Fam. Med.* 12, 568–572.
 Field, C.A., Welsh, A.H., 2007. Bootstrapping clustered data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 69, 369–390.
 Gold, R., Angier, H., Mangione-Smith, R., et al., 2012. Feasibility of evaluating the CHIPRA care quality measures in electronic health record data. *Pediatrics* 130, 139–149.
 Gold, R., Bailey, S.R., O'Malley, J.P., et al., 2014. Estimating demand for care after a Medicaid expansion: lessons from Oregon. *J. Ambul. Care Manage.* 37, 282–292.
 Gold, R., Burdick, T., Angier, H., et al., 2015. Improve synergy between health information exchange and electronic health records to increase rates of continuously insured patients. *eGEMS* 3, 1158.
 Gottlieb, L.M., Tirozzi, K.J., Manchanda, R., Burns, A.R., Sandel, M.T., 2015. Moving electronic medical records upstream: incorporating social determinants of health. *Am. J. Prev. Med.* 48, 215–218.
 Hall, J.D., Harding, R.L., DeVoe, J.E., et al., 2017. Designing health information technology tools to prevent gaps in public health insurance. *J. Innov. Health Inf.* 24, 900.
 Harding, R.L., Hall, J.D., DeVoe, J.E., et al., 2017. Maintaining public health insurance benefits: how primary care clinics help keep low-income patients insured. *Patient Experience J.* 4, 61–69.
 Hatch, B., Angier, H., Marino, M., et al., 2013. Using electronic health records to conduct children's health insurance surveillance. *Pediatrics* 132, e1584–e1591.
 Hatch, B., Tillotson, C., Angier, H., et al., 2016. Using the electronic health record for assessment of health insurance in community health centers. *J. Am. Med. Inform. Assoc.* 23, 984–990.
 Hatch, B., Marino, M., Killerby, M., et al., 2017. Medicaid's impact on chronic disease biomarkers: a cohort study of community health center patients. *J. Gen. Intern. Med.* 32, 940–947.
 Heintzman, J., Bailey, S.R., Hoopes, M.J., et al., 2014a. Agreement of Medicaid claims and electronic health records for assessing preventive care quality among adults. *J. Am. Med. Inform. Assoc.* 21, 720–724.
 Heintzman, J., Marino, M., Hoopes, M., et al., 2014b. Using electronic health record data to evaluate preventive service utilization among uninsured safety net patients. *Prev. Med.* 67, 306–310.
 Heintzman, J., Marino, M., Hoopes, M., et al., 2015. Supporting health insurance expansion: do electronic health records have valid insurance verification and enrollment data? *J. Am. Med. Inform. Assoc.* 22, 909–913.
 Heisey-Grove, D., Patel, V., 2015. Any, Certified, and Basic: Quantifying Physician EHR Adoption through 2014. The Office of the National Coordinator for Health Information Technology, Washington, D.C.
 Henry J. Kaiser Family Foundation, 2014. Children's Health Coverage: Medicaid, CHIP and the ACA. Menlo Park, CA.
 Hirsch, A.G., Scheck McAlearney, A., 2014. Measuring diabetes care performance using electronic health record data: the impact of diabetes definitions on performance measure outcomes. *Am. J. Med. Qual.* 29, 292–299.
 Hoopes, M.J., Angier, H., Gold, R., et al., 2016. Utilization of community health centers in Medicaid expansion and nonexpansion states, 2013–2014. *J. Ambul. Care Manage.* 39, 290–298.
 How is the ACA impacting Medicaid enrollment? <http://kff.org/medicaid/issue-brief/how-is-the-aca-impacting-medicaid-enrollment/>, Accessed date: 15 January 2015.
 HRSA awards \$7 million to new local health centers to help enroll people in the Health Insurance Marketplace. <https://www.hrsa.gov/about/news/press-releases/2015-11-13-outreach-enrollment.html>, Accessed date: 27 January 2018.
 Jones, E.B., Furukawa, M.F., 2014. Adoption and use of electronic health records among federally qualified health centers grew substantially during 2010–12. *Health Aff.* 33, 1254–1261.
 Kasper, J.D., Giovannini, T.A., Hoffman, C., 2000. Gaining and losing health insurance: strengthening the evidence for effects on access to care and health outcomes. *Med. Care Res. Rev.* 57, 298–318 (discussion 9–25).
 Kharrazi, H., Chi, W., Chang, H.Y., et al., 2017. Comparing population-based risk-stratification model performance using demographic, diagnosis and medication data extracted from outpatient electronic health records versus administrative claims. *Med. Care* 55, 789–796.
 Klompas, M., McVetta, J., Lazarus, R., et al., 2012. Integrating clinical practice and public health surveillance using electronic medical record systems. *Am. J. Prev. Med.* 42, S154–S162.
 Kottke, T.E., Baechler, C.J., Parker, E.D., 2012. Accuracy of heart disease prevalence estimated from claims data compared with an electronic health record. *Prev. Chronic Dis.* 9, E141.
 Lipsitz, S., Parzen, M., Fitzmaurice, G., Klar, N., 2003. A two-stage logistic regression

- model for analyzing inter-rater agreement. *Psychometrika* 68, 289–298.
- National Association of Community Health Centers, 2017. *Community Health Center Chartbook*.
- O'Malley, J.P., O'Keeffe-Rosetti, M., Lowe, R.A., et al., 2016. Health care utilization rates after Oregon's 2008 Medicaid expansion: within-group and between-group differences over time among new, returning, and continuously insured enrollees. *Med. Care* 54 (11), 984–991.
- Remington, P.L., Wadland, W.C., 2015. Connecting the dots: bridging patient and population health data systems. *Am. J. Prev. Med.* 48, 213–214.
- Soni, A., Hendryx, M., Simon, K., 2017. Medicaid expansion under the affordable care act and insurance coverage in rural and urban areas. *J. Rural. Health* 33, 217–226.
- Viera, A.J., Garrett, J.M., 2005. Understanding interobserver agreement: the kappa statistic. *Fam. Med.* 37, 360–363.