

# Supporting health insurance expansion: do electronic health records have valid insurance verification and enrollment data?

RECEIVED 15 November 2014  
 REVISED 20 January 2015  
 ACCEPTED 15 March 2015  
 PUBLISHED ONLINE FIRST 17 April 2015

John Heintzman<sup>1</sup>, Miguel Marino<sup>1</sup>, Megan Hoopes<sup>2</sup>, Steffani R Bailey<sup>1</sup>, Rachel Gold<sup>3</sup>, Jean O'Malley<sup>1</sup>, Heather Angier<sup>1</sup>, Christine Nelson<sup>2</sup>, Erika Cottrell<sup>2</sup>, Jennifer Devoe<sup>1</sup>



## ABSTRACT

**Objective** To validate electronic health record (EHR) insurance information for low-income pediatric patients at Oregon community health centers (CHCs), compared to reimbursement data and Medicaid coverage data.

**Materials and Methods** *Subjects* Children visiting any of 96 CHCs ( $N = 69\ 189$ ) from 2011 to 2012. *Analysis* The authors measured correspondence (whether or not the visit was covered by Medicaid) between EHR coverage data and (i) reimbursement data and (ii) coverage data from Medicaid.

**Results** Compared to reimbursement data and Medicaid coverage data, EHR coverage data had high agreement (87% and 95%, respectively), sensitivity (0.97 and 0.96), positive predictive value (0.88 and 0.98), but lower kappa statistics (0.32 and 0.49), specificity (0.27 and 0.60), and negative predictive value (0.66 and 0.45). These varied among clinics.

**Discussion/Conclusions** EHR coverage data for children had a high overall correspondence with Medicaid data and reimbursement data, suggesting that in some systems EHR data could be utilized to promote insurance stability in their patients. Future work should attempt to replicate these analyses in other settings.

**Keywords:** health insurance, electronic health records, Medicaid expansion, community health centers, health insurance claims, children

## INTRODUCTION

The Affordable Care Act (ACA) expanded health insurance options and mandated coverage for most Americans.<sup>1</sup> As these expansions are implemented, patients may increasingly seek coverage assistance from primary care clinics.<sup>2</sup> With recent expansions in electronic health record (EHR) adoption, EHR data and tools could help these clinics give patients health insurance enrollment and retention support.<sup>2–4</sup> Doing so will require clinics to have accurate and complete coverage information, so it is important to know whether patients' coverage is valid in EHR data available at the visit. The degree to which this information is available and accurate in EHR data is unknown.

We sought to develop a method for validating the EHR insurance coverage data seen in “real time” by clinics, and to use this method to understand availability and accuracy of health insurance data in a multi-site EHR. There is no established method for validating EHR insurance data, and the literature is unclear on whether reimbursement data (i.e., who actually paid for the visit) or payor enrollment data (i.e., who is actively enrolled in a plan) should be the “gold standard” source. Thus, we compared EHR insurance coverage data from an EHR shared across multiple primary care sites<sup>5–7</sup> to both (i) reimbursement data and (ii) payor coverage data on pediatric patients in Oregon. We hypothesized that EHR coverage information at a given visit would demonstrate good overall correspondence with the other two sources of insurance information. Findings from this study may support development of EHR-based tools that inform clinic staff about

patients' health insurance status, and engage staff and patients in ensuring insurance stability.

## METHODS

### Setting/Study Population

This retrospective validation study included the 96 Oregon clinics “live” on the Oregon Community Health Information Network (OCHIN) EHR by the study start date. OCHIN (previously the Oregon Community Health Information Network, now just OCHIN, as clinics from other states joined), is a national leader in the development of linked safety net EHRs. The OCHIN collaborative of primary care clinics share an Epic<sup>®</sup> EHR with data on >1.4 million patients in numerous states.<sup>5–7</sup> In addition to hosting a linked EHR, OCHIN supports healthcare innovation and has an active practice-based research network.<sup>5–7</sup> Our study population included children age <19 with  $\geq 1$  primary care visit in 2011–2012 ( $N = 76\ 147$  children). We chose a pediatric population because several policy initiatives during the study time period focused on improving children's coverage, yet many US children remain uninsured.<sup>8–11</sup> We excluded patients with private insurance, Medicare, or emergency Medicaid ( $n = 6416$ ), as we lacked full access to insurer data from these sources; pregnant teens ( $n = 530$ ) because of their unique public insurance options; and anyone who died during the study period ( $n = 12$ ). The final study population included  $N = 69\ 189$  pediatric patients with 287 846 visits in the study period.

Correspondence to John Heintzman MD, Oregon Health and Science University, Department of Family Medicine, 3181 SW Sam Jackson Park Rd, Portland, OR 97239, USA; heintzma@ohsu.edu; Tel: 503-494-2542; Fax: 503-346-6846

© The Author 2015. Published by Oxford University Press on behalf of the American Medical Informatics Association. All rights reserved.

For Permissions, please email: journals.permissions@oup.com

For numbered affiliations see end of article.

## Data Sources

We utilized three datasets, originating from two master sources (EHR and Medicaid data). The EHR data came from the OCHIN Epic<sup>®</sup> database, which includes comprehensive demographic, appointments, billing/reimbursement (including Medicaid unique client identification (ID) numbers), and clinical data. This was our source of demographic, utilization, and insurance coverage information from the date of each visit. We also obtained data from the OCHIN EHR regarding the payor who eventually paid for each visit (e.g., Medicaid, patient self-pay). Medicaid coverage data came from Oregon's Medicaid program. These three datasets and insurance coverage variables are described below.

1. *EHR coverage data.* This dataset contains dates of insurance coverage for clinic patients, contained in the EHR; these data reflect information available at the time of the visit, and are used to determine insurance coverage status at the time of the visit.
2. *Reimbursement data.* This dataset includes information regarding the payor who ultimately paid the claim(s) for a given clinic visit, which is recorded in the OCHIN EHR *after* a bill is paid. The eventual payor can differ from the original billed insurer if claims are denied, initial insurance information is incorrect, or a patient subsequently enrolls in an insurance program that covers services previously received.
3. *Medicaid coverage data.* This dataset, from the Oregon Medicaid program, contains date ranges of Medicaid coverage. These records were matched to patients and visit dates in the EHR datasets using Medicaid unique client ID numbers.

## Analysis:

We compared the EHR coverage data to the reimbursement data and the Medicaid coverage data. To examine variability in correspondence of data sources between the study clinics, we also performed these comparisons for each clinic. The primary outcome was dichotomous: whether or not the child had Medicaid insurance coverage at the time of visit.

We calculated common statistical measures of correspondence: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), agreement, and kappa statistic. *Sensitivity* is the probability that a dataset denoted coverage when the assumed gold standard did the same. *Specificity* is the proportion of encounters correctly classified as “no coverage” by the comparator data set, when the assumed gold standard denoted no coverage. PPV is the likelihood of a child being covered when the data set denoted coverage, and NPV is the likelihood of the child not being covered when the data set denoted no coverage. *Agreement* is defined as the total proportion of encounters in which the compared datasets denote the same coverage status. The *kappa statistic* is similar to agreement, but removes agreement that would be expected purely by chance. Cut-offs for good and excellent agreement/kappa statistic are 0.6 and 0.8, respectively.<sup>12</sup> Statistical analyses used SAS version 9.3 (SAS Institute, Inc.) and R version 3.1.0 (R Development Core Team). This study was approved by the Institutional Review Board at Oregon Health & Science University.

## RESULTS

### Patient Characteristics

Among the study population of 69 189 children, there were equal proportions by sex; 41% were aged 5–12 years; 44% were Hispanic, 40% non-Hispanic white; 57% spoke English; and 68% were in households meeting income criteria for Medicaid ( $\leq 100\%$  of the

Table 1: Characteristics of study patients, 2011–2012

	N (%)
No. of subjects	69 189 (100)
No. of encounters	287 846
Gender	
Female	34 380 (49.7)
Male	34 809 (50.3)
Age <sup>a</sup> , years	
<1	9495 (13.7)
1–4	15 798 (22.8)
5–12	28 273 (40.9)
13–1	15 623 (22.6)
Race/Ethnicity	
Hispanic	30 272 (43.8)
Non-Hispanic white	27 967 (40.4)
Non-Hispanic other	8505 (12.3)
Missing/unknown	2445 (3.5)
Language	
English	39 124 (56.6)
Spanish	23 118 (33.4)
Other	5020 (7.3)
Missing/unknown	1927 (2.8)
Household income <sup>b</sup>	
$\leq 100\%$ FPL	46 723 (67.5)
101–200% FPL	10 030 (14.5)
>200% FPL	6989 (10.1)
Missing/unknown	5447 (7.9)
Number of visits in study period	
1	19 131 (27.7)
2–3	21 671 (31.3)
4–5	11 648 (16.8)
6+	16 739 (24.2)

Data source: OCHIN EHR.

<sup>a</sup>Age assessed at earliest visit date in study period.

<sup>b</sup>Household income averaged across study period and presented as percent of federal poverty level (FPL); values  $\geq 1000\%$  FPL set to missing.

federal poverty level) (Table 1). There were 287 846 visits in the study period. Most patients had one to six visits; the median was three visits.

### Agreement between EHR Coverage Data and Reimbursement Data

All statistics for comparison are summarized in Table 2; raw data used to calculate these are included in Supplementary Appendix 1. The EHR coverage data and reimbursement data had high agreement (0.87),

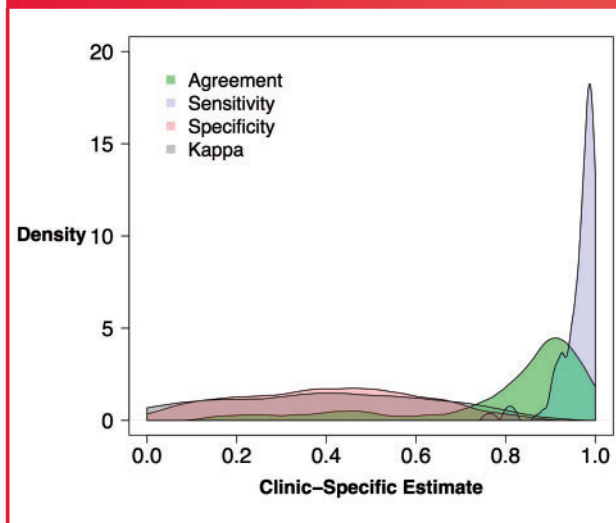
**Table 2: Pairwise Measures of Correspondence of Children's Coverage at Visits from 3 Data Sources: EHR Coverage, Reimbursement, and Medicaid Coverage**

Data sets being compared	Assumed Gold Standard	Agreement	Kappa Statistic	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
EHR Coverage	EHR Reimbursement	<b>86.6%</b>	<b>0.321</b> (0.316, 0.326)	<b>0.975</b> (0.974, 0.976)	<b>0.268</b> (0.264, 0.272)	<b>0.880</b> (0.879, 0.881)	<b>0.660</b> (0.653, 0.667)
EHR Coverage	Medicaid Coverage	<b>94.6%</b>	<b>0.486</b> (0.479, 0.493)	<b>0.964</b> (0.963, 0.965)	<b>0.595</b> (0.587, 0.603)	<b>0.980</b> (0.979, 0.980)	<b>0.453</b> (0.445, 0.460)

Bold numerical values represented estimated measures of correspondence and numbers in parentheses denote their corresponding 95% confidence interval.

**Figure 1: Distribution of Clinic-specific Correspondence Statistics in the Comparison of OCHIN EHR Coverage Data vs Reimbursement Data**

*Note:* Kernel Density Estimates of the distribution of clinic agreement, sensitivity, specificity, and kappa statistics. The kernel density estimator is a nonparametric method to estimate the probability distribution of the four statistical measures.



sensitivity (0.97), and PPV (0.88). When compared to the reimbursement data, the EHR coverage data had low kappa statistic (0.32), specificity (0.27), and moderate NPV (0.66). Notably, 11% of encounters classified as “covered” in the EHR were *not* documented in the reimbursement data as paid by Medicaid (see Appendix 1).

#### Agreement between EHR Coverage Data and Medicaid Coverage Data

Compared to Medicaid coverage data, EHR data had high agreement (0.95), sensitivity (0.96) and PPV (0.98), and lower kappa statistic (0.49), specificity (0.60), and NPV (0.45) (see Table 2).

#### Variability Among Clinics

Figure 1 shows the clinic-specific variability in agreement, kappa statistic, sensitivity, and specificity between EHR coverage and reimbursement data sources. Estimated clinic-specific sensitivity values of

EHR coverage and reimbursement data were generally high; 94% of clinics had sensitivity values  $>90\%$ . Agreement of coverage status did not vary substantially between clinics (Interquartile range: 77.6–93.1%). The distribution of kappa and specificity among all clinics is relatively flat and exhibited the most variation. The range of specificity values among all clinics was 0.06–0.76; for kappa statistic, 0.00–0.80. The other comparison (EHR vs. Medicaid coverage) had a similar distribution (data not shown).

#### DISCUSSION

Primary care clinics have improved care quality through the creation of patient-centered medical homes, care coordination, population management, patient engagement, and outreach.<sup>5,11,13–15</sup> Optimizing vulnerable populations' access to these services requires improving their access to stable health insurance coverage.<sup>8–10,13,16,17</sup> This study is significant because it 1) developed a feasible method for measuring the agreement of insurance information that could be adapted to other settings and 2) if replicated, could support insurance continuity for large numbers of Americans. Programs to actively assist patients with insurance enrollment and retention are not yet central to delivery system changes and are not usually supported by EHR-based tools.<sup>18–22</sup> In order to assist patients with health insurance enrollment and retention, clinics need valid insurance data in their EHRs.

To assess the extent to which insurance coverage data is available and valid at the time of a healthcare visit, we measured correspondence between EHR coverage, reimbursement, and Medicaid coverage data. We found that *EHR coverage information had high overall agreement with reimbursement and Medicaid data sets* for the pediatric population studied. Based on this study's findings, these clinics (and potentially other clinics with similar patient populations and EHRs) should consider using EHR data to help inform support staff, patients, and caregivers about current coverage and the potential need for a patient to re-enroll soon or to investigate new insurance coverage options available through the ACA. Given that insurance significantly predicts healthcare utilization and outcomes,<sup>23,24</sup> assistance with coverage enrollment and retention could be a potentially crucial service for clinics to provide<sup>2,3</sup> and could be facilitated by data and tools from the EHR. Building tracking systems to improve insurance coverage is arguably as (or more) important to patients' health as systems to improve blood pressure and lipid levels.<sup>3</sup> Our study presents validation for using the EHR as a source of health insurance information, and suggests that confidence in this information is possible. Related initiatives, now under way, will benefit greatly from these findings. For example, we have partnered with clinics to design and test tools that prompt staff when a patient's insurance coverage is lapsing<sup>2</sup>; knowing that the EHR data is a valid source of coverage information will help

propel this initiative. These analyses should be repeated to validate EHR insurance information in other networks to support similar work in those settings.

The pattern of our results (high agreement, sensitivity, and PPV along with lower specificity, NPV, and kappa statistics) suggests that EHR coverage data has a higher likelihood of correctly classifying insured visits than those reported as uninsured. For example, ~11% of visits where the EHR coverage data showed as covered were *not* shown to be paid by Medicaid in the reimbursement data. This may represent inaccurate insurance information, or alternative payment mechanisms (i.e., capitated payments) not captured in standard reimbursement data. It is possible that reimbursement data (which is analogous to claims data) may be less accurate in certain circumstances, especially if insurance coverage includes a per-member-per-month capitation algorithm or a similar global payment mechanism. These drawbacks to using payor data (i.e., claims reimbursed) could increase with primary care payment reform, if fee-for-service payments become less prevalent, and resultant claims data become less available. Our findings suggest that EHR coverage data may be comparable to, or better than, claims in identifying the insurance status of patients in low-income settings.

We saw significant variability in insurance information across study clinics, possibly due to differing workflows for obtaining insurance information or reimbursement. It is likely that these clinics have different payor mixes, which also could account for differences in kappa statistic. This highlights the need for improved clinic workflow practices and technologies for collecting accurate insurance information, and for developing systems to transfer this information directly from payors to healthcare clinics via EHR.

## LIMITATIONS

Although we studied data from nearly 100 clinics, our study was limited to one networked EHR and conducted in one state; other EHRs likely have different features and other states may have different Medicaid program rules. It is also uncertain if these results apply to adult populations. The methods we developed could be used to replicate these analyses in other EHRs, geographic regions, and populations, and with other insurance types, and should be repeated in other settings to have wider applicability. Additionally, as described above, we were limited in our ability to explain why ~11% of the patient visits had Medicaid coverage according to the EHR coverage data but were not paid by Medicaid, according to reimbursement data. Further studies using our methods could examine whether this pattern is seen in other systems and/or whether this discrepancy can be explained by alternate billing patterns (e.g., managed care payments, global capitation per-member-per-month). However, additional partnerships with insurers are needed to better understand this discrepancy. We also note that the kappa statistic was lower than the general agreement in this study. This can occur when one outcome is more prevalent (most of our studied encounters were “covered”),<sup>25</sup> thus suggesting that our high agreement scores may be partially due to the more common outcome (“coverage”), rather than whether the datasets actually corresponded well.

## CONCLUSION

We analyzed correspondence between EHR coverage data, reimbursement data, and Medicaid data on insurance coverage for children’s primary care visits. EHR coverage data had high agreement with the two other datasets. Our findings suggest that EHR systems’ coverage data is at least equal to Medicaid data in identifying insurance information for patients, and may even be more accurate. If replicated in

other studies, EHR data could be used to inform clinic staff and patients about insurance coverage status and could be incorporated into EHR-based tools aiming to improve the stability of insurance coverage for publicly insured Americans.

## FUNDING

This work was funded by the Patient Centered Outcomes Research Institute, Improving Health Systems (IMPACCT Kids’ Care, 2012); the National Cancer Institute (R01CA181452); and the Agency for Healthcare Quality and Research (K08HS021522).

## COMPETING INTEREST

None.

## CONTRIBUTORS

All listed authors meet ICMJE authorship criteria. J.H., M.M., M.H., S.B., and J.D. were involved in study design, data interpretation, and manuscript writing. M.H. and M.M. were responsible for data analysis as well. R.G., J.O.M., C.N., H.A., and E.C. were involved in data interpretation and manuscript writing.

## SUPPLEMENTARY MATERIAL

Supplementary material is available online at <http://jamia.oxfordjournals.org/>.

## REFERENCES

1. Kaiser Family Foundation. Summary of the Affordable Care Act 2013 [cited 2014]. <http://kff.org/health-reform/fact-sheet/summary-of-the-affordable-care-act/>. Accessed November 1, 2014.
2. DeVoe J, Angier H, Likumahwa S, et al. 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 Ambulatory Care Manag*. 2014;37(2):148–154.
3. DeVoe JE. Being uninsured is bad for your health: can medical homes play a role in treating the uninsurance ailment? *Ann Fam Med*. 2013;11(5):473–476.
4. DeVoe JA H, Burdick T, Gold R. Health information technology - an untapped resource for patient centered medical homes to help keep patients insured. *Ann Fam Med*. 2014;12(6):568–572.
5. 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. *JABFM*. 2013;26(3):271–278.
6. 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). *JABFM*. 2011;24(5):597–604.
7. OCHIN. Research and Analytics [cited 2014 June 5]. <https://ochin.org/offers/research/about-ochin-pbrn/>. Accessed November 1, 2014.
8. Newacheck PW, Stoddard JJ, Hughes DC, Pearl M. Health insurance and access to primary care for children. *New Engl J Med*. 338(8):513–519.
9. Stevens GD, Seid M, Halfon N. Enrolling vulnerable, uninsured but eligible children in public health insurance: association with health status and primary care access. *Pediatrics*. 1998;117(4):e751–e759.
10. Szilagyi PG, Holl JL, Rodewald LE, et al. Evaluation of New York State’s Child Health Plus: children who have asthma. *Pediatrics*. 2006;105(3 Suppl E):719–727.
11. Davidoff AJ, Garrett B. Determinants of public and private insurance enrollment among Medicaid-eligible children. *Med Care*. 2000;39(6):523–535.
12. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 37(5):360–363.

13. Baker LC, Afendulis C. Medicaid managed care and health care for children. *Health Services Res.* 2005;40(5 Pt 1):1466–1488.
14. Shi L, Frick KD, Lefkowitz B, Tillman J. Managed care and community health centers. *J Ambul Care Manag.* 2005;23(1):1–22.
15. Lebrun LA, Shi L, Chowdhury J, et al. Primary care and public health activities in select US health centers: documenting successes, barriers, and lessons learned. *Am J Public Health.* 2000;102 (Suppl 3):S383–S391.
16. Berman S, Armon C, Todd J. Impact of a decline in Colorado Medicaid managed care enrollment on access and quality of preventive primary care services. *Pediatrics.* 2005;116(6):1474–1479.
17. Damiano PC, Willard JC, Momany ET, Chowdhury J. The impact of the Iowa S-SCHIP program on access, health status, and the family environment. *Ambul Pediatr.* 2003;3(5):263–269.
18. Pourat N, Davis AC, Salce E, Hilberman D, Roby DH, Kominski GF. In ten California counties, notable progress in system integration within the safety net, although challenges remain. *Health Aff.* 2012;31(8):1717–1727.
19. Cunningham P, Felland L, Stark L. Safety-net providers in some US communities have increasingly embraced coordinated care models. *Health Aff.* 2012;31(8):1698–1707.
20. Denham AC, Hay SS, Steiner BD, Newton WP. Academic health centers and community health centers partnering to build a system of care for vulnerable patients: lessons from Carolina Health Net. *Acad Med.* 2013;88(5):638–643.
21. Neuhausen K, Grumbach K, Bazemore A, Phillips RL. Integrating community health centers into organized delivery systems can improve access to subspecialty care. *Health Aff.* 2012;31(8):1708–1716.
22. Katz MH, Brigham TM. Transforming a traditional safety net into a coordinated care system: lessons from healthy San Francisco. *Health Aff.* 2011;30(2):237–245.
23. Starfield B, Shi L. The medical home, access to care, and insurance: a review of evidence. *Pediatrics.* 2004;113(5 Suppl):1493–1498.
24. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Milbank Quart.* 2005;83(3):457–502.
25. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther.* 2005;85(3):257–268.

## AUTHOR AFFILIATIONS

<sup>1</sup>Oregon Health and Science University, Department of Family Medicine, Portland, OR 97239, USA

<sup>3</sup>Kaiser Center For Health Research Northwest, Portland, OR, USA

<sup>2</sup>OCHIN, Inc., Portland, OR, USA