

Calculations of Financial Incentives for Providers in a Pay-for-Performance Program

Manual Review Versus Data From Structured Fields in Electronic Health Records

Tracy H. Urech, MPH,* LeChauncy D. Woodard, MD, MPH,* Salim S. Virani, MD, PhD,*
R. Adams Dudley, MD, MBA,†‡ Meghan Z. Lutschg, BA,* and Laura A. Petersen, MD, MPH*

Background: Hospital report cards and financial incentives linked to performance require clinical data that are reliable, appropriate, timely, and cost-effective to process. Pay-for-performance plans are transitioning to automated electronic health record (EHR) data as an efficient method to generate data needed for these programs.

Objective: To determine how well data from automated processing of structured fields in the electronic health record (AP-EHR) reflect data from manual chart review and the impact of these data on performance rewards.

Research Design: Cross-sectional analysis of performance measures used in a cluster randomized trial assessing the impact of financial incentives on guideline-recommended care for hypertension.

Subjects: A total of 2840 patients with hypertension assigned to participating physicians at 12 Veterans Affairs hospital-based out-

patient clinics. Fifty-two physicians and 33 primary care personnel received incentive payments.

Measures: Overall, positive and negative agreement indices and Cohen's kappa were calculated for assessments of guideline-recommended antihypertensive medication use, blood pressure (BP) control, and appropriate response to uncontrolled BP. Pearson's correlation coefficient was used to assess how similar participants' calculated earnings were between the data sources.

Results: By manual chart review data, 72.3% of patients were considered to have received guideline-recommended antihypertensive medications compared with 65.0% by AP-EHR review ($\kappa=0.51$). Manual review indicated 69.5% of patients had controlled BP compared with 66.8% by AP-EHR review ($\kappa=0.87$). Compared with 52.2% of patients per the manual review, 39.8% received an appropriate response by AP-EHR review ($\kappa=0.28$). Participants' incentive payments calculated using the 2 methods were highly correlated ($r \geq 0.98$). Using the AP-EHR data to calculate earnings, participants' payment changes ranged from a decrease of \$91.00 (-30.3%) to an increase of \$18.20 ($+7.4\%$) for medication use (interquartile range, -14.4% to 0%) and a decrease of \$100.10 (-31.4%) to an increase of \$36.40 ($+15.4\%$) for BP control or appropriate response to uncontrolled BP (interquartile range, -11.9% to -6.1%).

Conclusions: Pay-for-performance plans that use only EHR data should carefully consider the measures and the structure of the EHR before data collection and financial incentive disbursement. For this study, we feel that a 10% difference in the total amount of incentive earnings disbursed based on AP-EHR data compared with manual review is acceptable given the time and resources required to abstract data from medical records.

Key Words: payment system design, performance measurement, incentives

(*Med Care* 2015;53: 901–907)

From the *Health Policy, Quality and Informatics Program, Michael E. DeBakey VA HSR&D Center for Innovations in Quality, Effectiveness & Safety (IQESt), and Section for Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX; †Department of Medicine, Philip R. Lee Institute for Health Policy Studies; and ‡Health Services Informatics, University of California, San Francisco, CA.

Preliminary findings were presented at the Society of General Internal Medicine 35th Annual Meeting in Orlando, FL (May 9–12, 2012) and the Academy Health Annual Research Meeting in Orlando, FL (June 24–26, 2012).

Supported in part by the US Department of Veterans Affairs, Veterans Health Administration, Health Services Research & Development (HSR&D) Investigator-Initiated Research (IIR) 04-349 (L.A.P.) and the National Institutes of Health (NIH) RO1 HL079173-01 (L.A.P.).

L.A.P. is the Director at the Center for Innovations in Quality, Effectiveness and Safety (CIN 13-413), Michael E. DeBakey VA Medical Center, Houston, TX. L.A.P. was a recipient of the American Heart Association Established Investigator Award (Grant number 0540043N). R.A.D. is a Robert Wood Johnson Investigator Awardee in Health Policy. The remaining authors declare no conflict of interest.

Reprints: Laura A. Petersen, MD, MPH, HSR&D (152), Michael E. DeBakey Veterans Affairs Medical Center, 2002 Holcombe Blvd, Houston, TX 77030. E-mail: laura.petersen@va.gov.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 0025-7079/15/5310-0901

Quality improvement initiatives, such as report cards and financial incentives linked to performance, must utilize clinical performance measures that are reliable, appropriate, timely, and cost-effective to process. Collecting these data by manual review is considered the gold standard but is very time and resource intensive. Electronic health records (EHRs) and structured data from electronic fields collected in

data warehouses are changing how health care providers document care and are increasingly viewed as a potential tool for performance measurement.

There is a nascent but growing body of literature comparing data derived by automated processing of structured fields in the electronic health record (AP-EHR) to measures collected by chart review. A study examining “meaningful use” clinical quality measures from electronic reporting and manual review found wide variation in the agreement for rates of recommended care for appropriate asthma medication, low-density lipoprotein cholesterol control for patients with diabetes mellitus, and pneumococcal vaccination.¹ A comparison of EHR-queried and chart review data from the New York City Primary Care Information Project found that EHR-derived measures under-reported patients meeting performance measures for 6 of 11 indicators.² Recently, Gardner et al³ reported low to moderate agreement for 12 measures of adolescent well-care across 3 sites when comparing data extracted from structured fields in the EHR to data collected by manual review.

Understanding how well data from structured fields in EHRs reflect the totality of what is known about the patient is critical as national quality reporting programs transition from manual review to automated reports, as they are largely doing so without natural language processing of free text as a supplement to measurement algorithms. The CMS EHR Incentive Program⁴ has spurred the implementation of EHR systems. In 2013, 58.9% of hospitals had a basic or comprehensive EHR, an increase from approximately 15% in 2010, the initial year of the CMS EHR incentive program.⁵ Starting in 2014, hospitals with certified EHR systems can submit 16 electronically specified clinical quality measures for the CMS Hospital Inpatient Quality Reporting program.⁶ Reporting methods for the CMS Physician Quality Reporting System, one of the pay-for-performance programs established under the Affordable Care Act, include direct EHR using certified EHR technology.⁷ New York City’s Health and Hospitals Corporation started a new pay-for-performance agreement linking physician bonus payments with quality and efficiency metrics, as well as implementation of an updated EMR system.^{8,9} Integrated Health Association’s pay-for-performance program, the largest non-governmental incentive program in the United States with approximately 35,000 physicians in California, uses only electronic data sources to assess clinical quality measures.¹⁰ These policy decisions notwithstanding, to our knowledge there have been no studies that have assessed the impact of basing payment on AP-EHR instead of chart review.

We sought to determine how well data from AP-EHR of structured fields from the Department of Veterans Affairs’ (VA) EHR reflect data collected by manual chart review and to assess the potential impact of the data collection methodologies on incentive earnings disbursed to physicians and practice groups participating in a trial evaluating pay-for-performance for hypertension care.

METHODS

Trial Design and Study Period

Primary care physicians and practice team members (eg, nurses) at 12 VA hospital-based outpatient clinics participated

in a trial evaluating the impact of financial incentives on guideline-recommended care for hypertension. The study’s design and main findings are described elsewhere.^{11,12} Briefly, hospitals were cluster randomized to 1 of 4 study groups—physician-level incentives, practice-level incentives, physician-level plus practice-level (combined) incentives, and no incentives (control). The intervention included 5 fourth-month performance periods. For each period, we randomly sampled and reviewed the EHRs for 40 eligible hypertensive patients from the physician’s panel. Participants in the intervention arms received incentive payments and all participants received audit and feedback performance reports.^{11,13} Participants earned payments based on the number of measures achieved per patient, with each measure rewarded the same amount.^{11,12} This study examines the trial’s fifth performance period, April through July 2009.

Hypertension Quality-of-Care Measures

We rewarded participants for prescribing guideline-recommended antihypertensive medications, achieving blood pressure (BP) thresholds, and appropriately responding to uncontrolled BP,¹¹ as described in the “Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure” (JNC 7).¹⁴ Appropriate responses to uncontrolled BP included prescribing lifestyle modifications for stage 1 hypertension or adjusting and/or starting guideline-recommended medications.

EHR Manual Record Review

Trained research assistants, blinded to the study’s objectives and study-group assignments, reviewed assigned health records in the Veterans Health Information Systems and Technology Architecture (Vista) Computerized Patient Record System (CPRS).^{15,16} They collected demographics; comorbidities; vitals including BP readings, heart rate, and pulse; names, dosages, and days supply for VA and non-VA prescribed medications; medication allergies, refusals, and contraindications; laboratory results; and lifestyle modifications.¹¹ Abstraction included reviewing the structured sections of the EHR and the free-text provider notes from the clinical encounters where hypertension should be addressed. Chart reviewers examined care delivery in many different outpatient settings, including but not limited to the primary care, cardiology, diabetes, renal/nephrology, emergency department, and women’s health. Some health information recorded during an encounter such as diagnoses and allergies/adverse reactions that are captured in structured sections can be linked to progress notes. Also, the abstraction tool contained questions to confirm the patient’s eligibility for the study, including verifying the patient had hypertension, was on the physician’s panel, and did not have a terminal illness.

AP-EHR Data Sources

We collected the same data elements listed in the chart abstraction tool from the AP-EHR of structured fields from the hospitals’ Vista/CPRS systems available in the VA Corporate Data Warehouse (CDW),¹⁷ the National Patient Care Database inpatient and outpatient SAS datasets,¹⁸ and

the VA Managerial Cost Accounting (MCA) Pharmacy and Laboratory National Data Extracts.¹⁹ These data sources capture all care provided in the VA system, not just at the 12 study sites. We also included data from the non-VA medical care files (care received outside the VA and paid by the VA)²⁰ and the VA-CMS merged datasets.²¹ VA health care providers (eg, physicians and nurses) enter their own data in the patient's medical record in CPRS.

Index Visit

Chart reviewers determined whether a patient had a qualifying outpatient visit during the performance period, defined as an encounter where hypertension should be addressed. We identified visits where BP readings were recorded using the CDW vital signs domain.

Use of Guideline-recommended Antihypertensive Medications

To determine whether patients were receiving guideline-recommended antihypertensive medications at index, we assessed their medications, accounting for coexisting conditions and allergies, contraindications, and refusals. On the basis of JNC 7 guidelines, we developed sophisticated medication scenario algorithms to determine medication appropriateness (Supplemental Appendix Table 1, Supplemental Digital Content 1, <http://links.lww.com/MLR/B7>). We identified patients' past and current medications using the MCA pharmacy data. Antihypertensive medications filled within 120 days before and including the index visit date were assessed. Although the standard prescription fill for these medications is 90 days, we included extra time for patients to refill medications or finish any stockpiled medications.

Using inpatient and outpatient diagnosis and procedure codes, we searched the AP-EHR data to identify patients' JNC 7 compelling indications and other conditions that impact pharmacologic treatment, applying the same time intervals used in the abstraction tool (Supplemental Appendix Table 2, Supplemental Digital Content 2, <http://links.lww.com/MLR/B8>). An internist from the trial study team and a cardiologist independently reviewed the codes.

To account for contraindications, allergies, refusals, and use of non-VA antihypertensive medications, we incorporated data from the allergy, vital signs, and health factors domains in the VA CDW. Providers are able to document refusals and use of non-VA antihypertensive medications in the free-text area of the Notes section in CPRS or by resolving a clinical reminder that creates text. Clinical reminder software in CPRS includes health factors, computerized components capturing patient information that does not have existing standard coding (eg, outside medication—ACE inhibitor).²² The clinical reminder software, when activated by the provider, uses dialog boxes to prompt the provider to check a box from a list of options that best reflects the patient encounter. Also, providers are able to document non-VA medications in structured fields in the Orders section in CPRS.

BP Control and Appropriate Response to Uncontrolled BP

We used outpatient BP readings from the VA CDW vital signs domain to assess each patient's BP control at index. We included BP readings from many types of clinics, including but not limited to primary care, cardiology, diabetes, renal/nephrology, emergency department, and women's health. Unlike the EHR manual review, we were not able to examine home or outside BP readings.

We used the MCA pharmacy data to examine anti-hypertensive medication adjustments in response to uncontrolled BP. Implementing the same approach as the EHR-abstracted data, we assessed appropriate responses at the index visit and during the follow-up period. We compared the patient's medications before the index visit, at index, and after the index visit. Using the prescription fill and daily dosage information, we identified dosage changes to existing medications. Medication adjustments were classified as either an increase, decrease, or no change in dosage.

We identified lifestyle modifications using outpatient visits and health factors. Using the list of modifications in the abstraction tool, we created the following categories to evaluate the AP-EHR data: nutrition/diet education and counseling, general patient education, exercise programs such as the Managing Obesity for Veterans Everywhere! (MOVE!), BP education visits, and pharmacist-led education and counseling. We gave credit if the lifestyle modification health factor indicated the patient was referred to one of the above programs or if the patient declined the referral. Providers are able to document lifestyle modifications in the free-text area of the Notes section in CPRS or by resolving a clinical reminder (eg, weight and nutrition screen).

Analysis

We calculated the overall, positive and negative agreement indices and Cohen's kappa for the JNC 7 compelling indications, other coexisting conditions, and the 3 performance measures.^{23–25} Overall agreement is the number of outcomes both data sources agree on divided by the total number of outcomes assessed. Positive/negative agreement is calculated by taking the number of positive/negative outcomes both data sources agree on and dividing by the number of positive/negative outcome identified by each data source. Cohen's kappa is a measure of agreement that corrects for agreement due to chance.²⁶ We used the following guidelines to interpret the kappa values: 0.01–0.20 indicates slight agreement; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; and 0.81–0.99 reflects almost perfect agreement.²⁷ Analyses were performed using SAS version 9.1 (Cary, NC).

We calculated the financial incentive payments for the intervention participants using the same approach from the trial. For the guideline-recommended medication measure and the combined measure assessing controlled BP or an appropriate response to uncontrolled BP, we compared the overall amount of money that would have disbursed using the AP-EHR data to the amount that was disbursed based on manual chart abstraction. We assessed how similar the participants' calculated earnings were between the 2 data collection methods using Pearson's correlation coefficient.

We also examined the payment changes when AP-EHR data was used to determine earnings.

This study was approved by Baylor College of Medicine Institutional Review Board and the Michael E. DeBakey Research and Development Committee. The funding sources had no role in the study.

RESULTS

We compared the performance measures for the randomly sampled 2840 patients assigned to the 71 physicians (52 in the incentive groups and 19 in the control group) participating in the fifth (and final) intervention period of the trial. For the JNC 7 compelling indications, we found almost perfect agreement for diabetes mellitus and chronic kidney disease between the manual review and the AP-EHR data (Table 1). Agreement was lowest for unstable angina.

In the manual review data, 2054 (72.3%) patients received guideline-recommended antihypertensive medications compared with 1847 (65.0%) patients in the AP-EHR data (Table 2). Overall agreement was 0.79 and kappa was 0.51 [95% confidence interval (CI), 0.48–0.55], indicating moderate agreement. As medication appropriateness was dependent on the compelling indications, we analyzed the subset of those patients who had the same compelling indication classification in both data sources ($n=2403$). Agreement increased slightly ($\kappa=0.56$; 95% CI, 0.53–0.60). When we included VA/CMS data to identify JNC 7 compelling indications not recorded in the AP-EHR data sources, agreement for medication use did not improve ($\kappa=0.50$; 95% CI, 0.46–0.53).

To assess the drivers of disagreement between the AP-EHR and manual review data, we reexamined 252 patient records that had VA antihypertensive medication use in both data sources and the same compelling indicator classification but did not agree on the performance measure. These represent 42.2% of the records that did not agree between the data sources. Manual review and AP-EHR data disagreed on only 1 element (eg, beta blocker allergy) in the medication algorithms for 128 (50.8%) records and only 2 elements for 67 (26.6%) records. Among these 262 disagreements, 235 (89.7%) were due to the element not being captured in

the AP-EHR data. Medication algorithm elements with the highest prevalence of disagreement were being prescribed a guideline-recommended thiazide medication ($n=56$, 21.4%), having a thiazide allergy/refusal ($n=51$, 19.5%), and having coexisting benign prostatic hyperplasia ($n=49$, 18.7%). Manual review documentation indicated that thiazide allergy/refusal and benign prostatic hyperplasia were primarily documented only in the Notes section of the record.

In the final performance period of the trial, 1974 (69.5%) patients had controlled BP based upon the manual review data. The proportion of patients meeting this measure was slightly lower (66.8%) using the AP-EHR data (Table 2). We report agreement indices separately for controlled BP and appropriate response to uncontrolled BP as appropriate response agreement could only be assessed for the subset of patients who had uncontrolled BP in both data sources ($n=825$). Among the subset, 52.2% received an appropriate response in the manual review data compared with 39.8% in the AP-EHR data. Of the 3 performance measures, agreement was highest for controlled BP and lowest for appropriate response, kappa values 0.87 (95% CI, 0.85–0.89) and 0.28 (95% CI, 0.22–0.35), respectively (Table 2). Agreement for controlled BP increased slightly after excluding the patients with controlled readings based on home/outside readings ($\kappa=0.90$; 95% CI, 0.88–0.92).

Among the 825 patients with uncontrolled BP in both data sources, 371 (45.0%) patients had documented lifestyle modifications according to the manual review data compared with 186 (22.6%) in the AP-EHR data. Overall agreement for lifestyle modifications was 0.60 and kappa was 0.16 (95% CI, 0.10–0.22). Restricting the analysis to the patients who had stage 1 hypertension in both data sources ($n=689$) (and for whom lifestyle modifications were recommended by guidelines), overall agreement for the appropriate response measure was 0.61 and kappa was 0.24 (95% CI, 0.17–0.31).

Fifty-two physicians and 33 nonphysician practice team members received incentive payments for the 4-month performance period. There were a total of 5 payment disbursements in the trial. Using manual review data, mean payments for the fifth performance period for the medication and the combined controlled BP or appropriate response to uncontrolled BP measures were \$212.40 (SD=\$124.61) and

TABLE 1. Agreement Between Manual Review and AP-EHR Data for Comorbid Conditions Examined to Evaluate Guideline-recommended Antihypertensive Medication Use

JNC 7 Compelling Indication or Comorbid Condition	No. Patients (%), Manual Review Data*	No. Patients (%), AP-EHR Data*	Overall Agreement	Positive Agreement	Negative Agreement	Kappa
Diabetes mellitus	1090 (38.4)	1056 (37.2)	0.96	0.95	0.97	0.92
Chronic kidney disease	705 (24.8)	611 (21.5)	0.96	0.92	0.98	0.90
Nephropathy	212 (7.5)	174 (6.1)	0.98	0.87	0.99	0.86
Ischemic heart disease	686 (24.2)	682 (24.0)	0.92	0.84	0.95	0.79
Myocardial infarction	262 (9.2)	157 (5.5)	0.94	0.56	0.97	0.53
Unstable angina	45 (1.6)	42 (1.5)	0.98	0.39	0.99	0.38
Benign prostatic hyperplasia	900 (31.7)	473 (16.7)	0.83	0.66	0.89	0.56
Angioedema	21 (0.7)	19 (0.7)	0.99	0.55	0.99	0.55
Cor pulmonale	20 (0.7)	0	—	—	—	—

*Examined 2840 patient records.

AP-EHR indicates automated processing of structured fields in the electronic health record; JNC 7, Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure.

TABLE 2. Agreement for Use of Guideline-recommended Antihypertensive Medications, Blood Pressure Control, and Appropriate Response to Uncontrolled Blood Pressure

Rewarded Clinical Measure	No. Patients (%), Manual Review Data	No. Patients (%), AP-EHR Data	Overall Agreement	Positive Agreement	Negative Agreement	Kappa
Use of guideline-recommended medications*	2054 (72.3)	1847 (65.0)	0.79	0.85	0.66	0.51
Controlled BP*	1974 (69.5)	1898 (66.8)	0.94	0.96	0.91	0.87
Appropriate response to uncontrolled BP†	431 (52.2)	328 (39.8)	0.64	0.61	0.66	0.28

*Examined 2840 patient records.

†Among the 825 patients who had uncontrolled blood pressure in both data sources.

AP-EHR indicates automated processing of structured fields in the electronic health record; BP, blood pressure.

\$246.34 (SD=\$126.60), respectively. The total amounts of incentives disbursed for guideline-recommended medications and controlled BP or appropriate response to uncontrolled BP were \$18,054.40 and \$20,939.10, respectively. Using the AP-EHR data, the participants' mean payments would have been \$191.21 (SD=\$111.42) for the medication measure and \$225.79 (SD=\$114.55) for controlled BP or appropriate response. Total incentive earnings would have decreased by 10.0% and 8.3%, respectively. Individual participants' incentive payments calculated using the 2 methods were highly correlated ($r \geq 0.98$). Using the AP-EHR data to calculate incentive earnings, payment changes ranged from a decrease of \$91.00 (−30.3%) to an increase of \$18.20 (+7.4%) for guideline-recommended medication use [interquartile range (IQR), −14.4% to 0%]. Payments would have decreased by 10% or more for 28 physicians and 11 practice team members (Fig. 1). For BP control or appropriate response to uncontrolled BP, changes ranged from a decrease of \$100.10 (−31.4%) to an increase of \$36.40 (+15.4%) (IQR, −11.9% to −6.1%). Payments would have decreased by 10% or more for 18 physicians and 9 practice team members.

DISCUSSION

In this assessment of how well the AP-EHR reflect data from manual review and the impact of these data on performance rewards, we found almost perfect agreement for the BP control measure and low agreement for the appropriate response to uncontrolled BP measure. We also found moderate agreement between the data sources for use of guideline-recommended antihypertensive medication, a measure requiring information about the patients' coexisting conditions, current medications, and medications allergies, contraindications, and refusals. We determined that the total amount of incentives disbursed to physicians and practice team members would have been lower using the AP-EHR data to reward performance. Using the trial's reward structure of individual-level, practice-level, and combined incentives, we found that payments calculated using the manual review and AP-EHR data were highly correlated.

Considering the manual review as the gold standard, the AP-EHR data underreported the number of patients receiving appropriate medications. This finding is similar to other studies comparing EHR-derived data and manual review.^{1,28} For BP agreement, we used the vital signs domain

from the CDW. Borzecki et al²⁹ found that these data from VA CPRS are sufficient for assessing the quality of hypertension care and reviewing clinicians' notes in the EHR for BP readings adds little information.

This study's novel contributions to the literature are 2-fold. First, we are not aware of any studies that have compared manual review data to data from automated sources to evaluate providers' responses to uncontrolled BP. Secondly, and more importantly, we are not aware of any studies that have assessed how the differences in performance measures from manually reviewed data and EHR automated data impact providers financially in pay-for-performance plans. We feel that a 10% difference in the total amount of incentive earnings disbursed based on AP-EHR data compared with manual review is acceptable given the time and resources required to abstract data from medical records. Of the 71 physicians examined, 46 (66.2%) and 56 (78.9%) had a percent change in earnings <10%, for the use of guideline-recommended antihypertensive medications

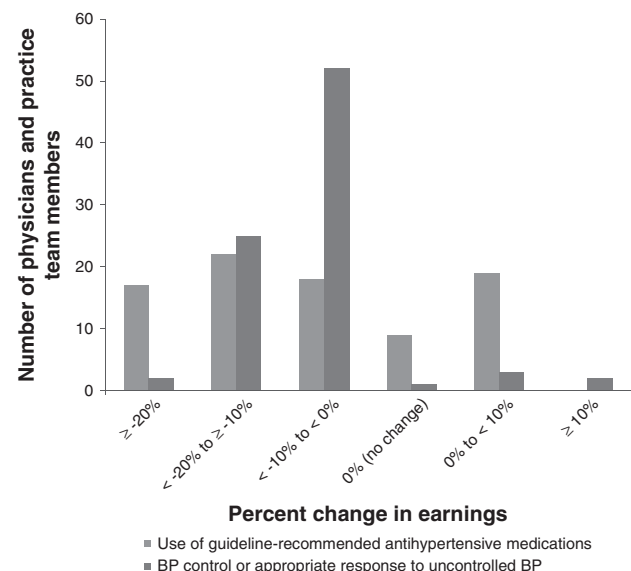


FIGURE 1. Percent change in participants' performance incentive payments using AP-EHR data sources. AP-EHR indicates automated processing of structured fields in the electronic health record; BP, blood pressure.

measure and the BP control or appropriate response to uncontrolled BP measure, respectively.

Our data are consistent with several studies that suggest that discrepancies between EHR-derived data and manual review are due to information documented in unstructured fields or as unstructured text in the health record.^{1–3,30,31} Understanding that the AP-EHR data sources are not a perfect reflection of the information in the patient's health record, we were not surprised that only 64.2% patients were identified as receiving guideline-recommended medications compared with 72.3% per the manual review. We assessed up to 19 elements in the medication algorithms for the compelling indication classifications, including coexisting conditions, antihypertensive medication classes, and allergies, contraindications, and refusals. Because the medication algorithm requires many pieces of clinical information that may not be documented in structured fields, it is more susceptible to missing data. We found in our reexamination of a subset of records that 2 of the 3 most prevalent disagreed elements in the medication algorithms were more often captured only in the Notes section of the patient's electronic medical record. If the trial had collected data by the AP-EHR approach instead of manual record review, our findings might have been different given that providers might have been more likely to document clinical information and interactions in structured fields in the EHR.

Our study has several limitations. First, we compared agreement for performance measures and the incentive payments for the records the reviewers identified as eligible. Manual review found that approximately 20% of the patient records identified from the physicians' sampling frames developed using AP-EHR data were not eligible for the study. If our study used only AP-EHR data sources, we would have provided performance audit and feedback and financial rewards that included inappropriate patients, unfairly characterizing the physician's and practice's performance. Second, we were not able to explore why there was low agreement for the appropriate response to uncontrolled BP measure. Most providers received credit for this measure because a lifestyle modification was recommended. We hypothesize that this information is documented in the Notes section of the record, but we were not able to systematically evaluate this. Third, given the VA health care system is a unique environment with a national EHR and a performance measurement reporting system, VA providers are more likely to be aware of documenting clinical information electronically compared with those delivering care within independent small and medium-sized physician practices. Although the unique organizational culture of the VA along with its predominantly male and older Veteran population and the salaried (vs. fee-for-service) providers impact the generalizability of our findings, the multiple electronic data sources in the VA, especially the VA CDW, reflect current initiatives in health care systems, both at the national and state levels, to use EHR-derived data to evaluate clinical performance and efficiency.

Pay-for-performance plans that use only EHR-derived data to reward providers should carefully consider the measures and the structure of the EHR and other electronic

data sources before data collection and financial incentive disbursement. Quality assessment that includes metrics that require several different data elements, such as guideline-recommended medications, and actions that occur during the patient-provider encounter that are not recorded in a structured field are more likely to underestimate provider performance.

ACKNOWLEDGMENTS

The authors acknowledge Mark Kuebler, MS, for his programming and data analysis, Kate Simpson Ramos, MPH, for her contribution to the methods, and Christine Yang, BA, for her assistance with the literature review for this manuscript.

REFERENCES

1. Kern LM, Malhotra S, Barron Y, et al. Accuracy of electronically reported "meaningful use" clinical quality measures. *Ann Intern Med.* 2013;158:77–83.
2. Parsons A, McCullough C, Wang J, et al. Validity of electronic health record-derived quality measurement for performance monitoring. *J Am Med Inform Assoc.* 2012;19:604–609.
3. Gardner W, Morton S, Byron SC, et al. Using computer-extracted data from electronic health records to measure the quality of adolescent well-care. *Health Serv Res.* 2014;49:1226–1248.
4. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare and Medicaid programs: hospital outpatient prospective payment and ambulatory surgical center payment systems and quality reporting programs; Hospital Value-Based Purchasing Program; organ procurement organizations; quality improvement organizations; Electronic Health Records (EHR) Incentive Program; provider reimbursement determinations and appeals. Final rule with comment period and final rules. *Fed Regist.* 2013;78:74825–75200.
5. Adler-Milstein J, DesRoches CM, Furukawa MF, et al. More than half of US hospitals have at least a basic EHR, but stage 2 criteria remain challenging for most. *Health Aff (Millwood).* 2014;33:1664–1671.
6. Electronically specified clinical quality measures (eCQMs) reporting quality net. Available at: <http://www.qualitynet.org>. Accessed March 25, 2015.
7. Centers for Medicare and Medicaid Services. Centers for Medicare and Medicaid Services Physician Quality Reporting System (PQRS) Overview. Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/PQRS_OverviewFactSheet_2013_08_06.pdf. Accessed June 23, 2015.
8. Hartocollis A. New York City ties doctors' income to quality of care. *New York Times.* January 12, 2013:A1. Available at: <http://www.nytimes.com/2013/01/12/nyregion/new-york-city-hospitals-to-tie-doctors-performance-pay-to-quality-measures.html?pagewanted=all&r=2&>. Accessed October 3, 2014.
9. New York City Health and Hospitals Corporation. HHC Signs Contract for New Electronic Medical Record System to Span City's Entire Public Hospital System (press release, dated January 16, 2013). Available at: <http://www.nyc.gov/html/hhc/html/news/press-release-2013-01-16-medical-records-contract.shtml>. Accessed June 23, 2015.
10. Integrated Health Association (IHA). Integrated Health Association (IHA) Pay for Performance (P4P) Program Fact Sheet. Available at: http://www.ih.org/pdfs_documents/p4p_california/P4P-Fact-Sheet-September-2014.pdf. Accessed October 3, 2014.
11. Petersen LA, Urech T, Simpson K, et al. Design, rationale, and baseline characteristics of a cluster randomized controlled trial of pay for performance for hypertension treatment: study protocol. *Implement Sci.* 2011;6:114.
12. Petersen LA, Simpson K, Pietz K, et al. Effects of individual physician-level and practice-level financial incentives on hypertension care: a randomized trial. *JAMA.* 2013;310:1042–1050.
13. Hysong SJ, Simpson K, Pietz K, et al. Financial incentives and physician commitment to guideline-recommended hypertension management. *Am J Manage Care.* 2012;18:e378–e391.

14. Chobanian AV, Bakris GL, Black HR, et al. National Heart, Lung, and Blood Institute Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; National High Blood Pressure Education Program Coordinating Committee. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*. 2003;289:2560–2572.
15. Brown SH, Lincoln MJ, Groen PJ, et al. VistA—US Department of Veterans Affairs national-scale HIS. *Int J Med Inform*. 2003;69:135–156.
16. Department of Veterans Affairs Office of Enterprise Development. VistA-HealtheVet Monograph. Available at: http://www.va.gov/vista_monograph/docs/2008_2009_vistahealthevet_monograph_fc_0309.pdf. Accessed August 28, 2014.
17. Fihn SD, Francis J, Clancy C, et al. Insights from advanced analytics at the Veterans Health Administration. *Health Aff*. 2014;33:1203–1211.
18. Hynes DM, Perrin RA, Rappaport S, et al. Informatics resources to support health care quality improvement in the Veterans Health Administration. *J Am Med Inform Assoc*. 2004;11:344–350.
19. Smith MW, Joseph GJ. Pharmacy data in the VA health care system. *Med Care Res Rev*. 2003;60:92S–123S.
20. Department of Veterans Affairs Health Economics Resource Center (HERC). Non-VA medical care (formerly fee basis files). Available at: <http://www.herc.research.va.gov/include/page.asp?id=non-va>. Accessed March 25, 2015.
21. Department of Veterans Affairs. VHA Directive 2010-019: access to Centers for Medicare and Medicaid Services (CMS) data for VHA users within the VA information technology (IT) systems. Available at: http://www.va.gov/vhapublications/ViewPublication.asp?pub_ID=2228. Accessed September 15, 2014.
22. Department of Veterans Affairs. VHA Directive 2008-084: National Clinical Reminders. Available at: http://www.va.gov/vhapublications/ViewPublication.asp?pub_ID=1815. Accessed September 3, 2014.
23. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol*. 1990;43:551–558.
24. Chen G, Faris P, Hemmelgarn B, et al. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Med Res Methodol*. 2009;9:5.
25. Barbara AM, Loeb M, Dolovich L, et al. Patient self-report and medical records: Measuring agreement for binary data. *Can Fam Physician*. 2011;57:737–738.
26. Cohen J. A coefficient of agreement for nominal scales. *Edu and Psychol Meas*. 1960;20:37–46.
27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
28. Persell SD, Wright JM, Thompson JA, et al. Assessing the validity of national quality measures for coronary artery disease using an electronic health record. *Arch Intern Med*. 2006;166:2272–2277.
29. Borzecki AM, Wong AT, Hickey EC, et al. Can we use automated data to assess quality of hypertension care? *Am J Manag Care*. 2004;10:473–479.
30. 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;146:270–277.
31. Garrido T, Kumar S, Lekas J, et al. e-Measures: insight into the challenges and opportunities of automating publicly reported quality measures. *J Am Med Inform Assoc*. 2014;21:181–184.