Validation of Stroke Meaningful Use Measures in a National Electronic Health Record System

Michael S. Phipps, MD, MHS^{1,2}, Jeff Fahner, BSN, MS³, Danielle Sager, BS³, Jessica Coffing, MPH³, Bailey Maryfield, MS³, and Linda S. Williams, MD^{3,4,5}

¹Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA; ²Baltimore VA Medical Center, Baltimore, MD, USA; ³Roudebush VA Medical Center, Indianapolis, IN, Indiana; ⁴Department of Neurology, Indiana University School of Medicine, Indianapolis, IN, USA; ⁵Regenstrief Institute Inc., Indianapolis, IN, USA.

BACKGROUND: The Meaningful Use (MU) program has increased the national emphasis on electronic measurement of hospital quality.

OBJECTIVE: To evaluate stroke MU and one VHA stroke electronic clinical quality measure (eCQM) in national VHA data and determine sources of error in using centralized electronic health record (EHR) data.

DESIGN: Our study is a retrospective cross-sectional study of stroke quality measure eCQMs vs. chart review in a national EHR. We developed local SQL algorithms to generate the eCQMs, then modified them to run on VHA Central Data Warehouse (CDW) data. eCQM results were generated from CDW data in 2130 ischemic stroke admissions in 11 VHA hospitals. Local and CDW results were compared to chart review.

MAIN MEASURES: We calculated the raw proportion of matching cases, sensitivity/specificity, and positive/negative predictive values (PPV/NPV) for the numerators and denominators of each eCQM. To assess overall agreement for each eCQM, we calculated a weighted kappa and prevalence-adjusted bias-adjusted kappa statistic for a three-level outcome: ineligible, eligible-passed, or eligible-failed.

KEY RESULTS: In five eCQMs, the proportion of matched cases between CDW and chart ranged from 95.4 %–99.7 % (denominators) and 87.7 %–97.9 % (numerators). PPVs tended to be higher (range 96.8 %–100 % in CDW) with NPVs less stable and lower. Prevalence-adjusted bias-adjusted kappas for overall agreement ranged from 0.73–0.95. Common errors included difficulty in identifying: (1) mechanical VTE prophylaxis devices, (2) hospice and other specific discharge disposition, and (3) contraindications to receiving care processes.

CONCLUSIONS: Stroke MU indicators can be relatively accurately generated from existing EHR systems (nearly 90 % match to chart review), but accuracy decreases slightly in central compared to local data sources. To improve stroke MU measure accuracy, EHRs should include standardized data elements for devices, discharge disposition (including hospice and comfort care status), and recording contraindications.

KEY WORDS: stroke; meaningful use; electronic health records; quality assessment; process assessment.

J Gen Intern Med 31(Suppl 1):S46–52

DOI: 10.1007/s11606-015-3562-5

© Society of General Internal Medicine 2015

INTRODUCTION

Quality measurement use for ischemic stroke in the US has expanded considerably in the past decade.^{1,2} The Joint Commission requires eight quality measures in its Primary Stroke Center certification program, those participating in the Get With the Guidelines program from the American Stroke Association need to report these eight and other quality measures, and the Centers for Medicare and Medicaid now publicly report ischemic stroke mortality and readmissions on their Hospital Compare website.³ The Affordable Care Act included the eight stroke indicators in its Meaningful Use (MU) program,⁴ which provides incentives to institutions whose electronic health records (EHRs) are able to automatically collect and report hospital-based quality measures without manual chart review. It is unclear at this time how many EHRs are able to meet this standard, and the extent and nature of errors in the collection and reporting of these data are not known.

The Veterans Health Administration (VHA) does not participate in the MU EHR incentive program, but complies with other national quality measurement programs (e.g., HospitalCompare.gov) and has a very long history of data-driven quality management. In 2012, the VHA adopted three self-reported stroke quality indicators and in January 2015 started to conduct External Peer Review Program (EPRP) chart reviews for the eight inpatient stroke MU measures, a process that involves contracted abstractors manually reviewing charts 1 to 2 months after discharge.⁵ This process maximizes accuracy but can be costly and time consuming, and it often results in a delay between patient care and reporting of these measures. There is increasing interest in using electronic clinical quality measures (eCQMs) not only to reduce the cost of obtaining data but also to provide a closer pairing of the quality feedback to the time of patient care.

Electronic supplementary material The online version of this article (doi:10.1007/s11606-015-3562-5) contains supplementary material, which is available to authorized users.

Published online March 7, 2016

The purpose of this study was to develop and validate inpatient stroke eCQMs that are part of the MU program and are relevant to the VHA's desire to monitor and improve inpatient stroke care. This project examined four stroke MU measures and one VHA stroke eCQM in national VHA data and determined sources of error in the eCQMs compared to standardized chart review.

METHODS

Overview

This project was conducted as an operational initiative in partnership between the Indianapolis Center for Health Information and Communication and the VHA Office of Analytics and Business Intelligence (OABI). Project activities took place under a jointly approved Memorandum of Understanding between the Richard L. Roudebush VA Medical Center (RVAMC) and the OABI. Collection of the criterion standard chart review data took place as part of a prior research project, the Intervention for Stroke Performance Improvement using Redesign Engineering (INSPIRE) project;⁶ use of these prior chart review data for the purpose of validating the eCQMs was approved by the local IRB and VA R&D committees.

We used multiple sources of VHA data, including data from the local Veterans Health Information Systems and Technology (VistA) EHR files (which include individual patient record data in both nationally standardized and locally adapted data elements including 'health factors' and 'orderable items') and from the CDW (see [Supplemental Appendix](#) for a complete list of data elements used).⁷ VistA data from every VHA facility are updated every 24 h into the CDW, which also incorporates data from other VHA data systems. CDW data are stored in relational databases and organized into discrete data domains. Multiple types of patient identifiers allow identification of unique patients and linkage of individual patient data across CDW data tables. For this study, we used patient social security numbers and admission/discharge dates to match CDW data to the identified and chart review verified stroke admissions from the INSPIRE study.

Measure Development

We initially constructed Structured Query Language (SQL) queries to generate five eCQMs: four MU measures following the specifications for the CMS Stroke National Inpatient Quality Measures version 4.2a⁸ effective for calendar year 2013 and the VHA NIH Stroke Scale (NIHSS) indicator following the VHA Inpatient Evaluation Center (IPEC) specifications for this indicator. MU measures developed included: (1) STK-1, venous thromboembolism (VTE) prophylaxis; (2) STK-2: antithrombotic (AT) at discharge; (3) STK-5: AT by hospital day 2 (AT by HD2); (4) STK-10: consider for rehabilitation (Rehab). The SQL queries initially were developed to run on local VistA EHR data housed in the VISN 11 Regional Data

Warehouse. We used text mining within the SQL queries to identify NIHSS performance in any notes and to assess for mentions of "hospice" in the discharge summary only (see [Supplemental Appendix](#) for indicator algorithms and data definitions). For each eCQM, we compared the SQL algorithm results to the results of local chart review on all confirmed ischemic stroke admissions to the RVAMC in the year 2013 (N = 98), identified from discharge ICD-9 codes as specified in the stroke national inpatient quality measure specifications.⁸ Although they are included in the national specifications, we did not include hemorrhagic stroke admissions because the criterion standard chart review cases from the INSPIRE study included only ischemic stroke admissions. Hemorrhagic stroke admissions are a small percentage of all strokes (about 10–15 %); therefore, we do not expect this exclusion to dramatically affect results.

Criterion Standard

We compared each eCQM denominator and numerator to the corresponding chart review indicator result. Chart review indicators were assessed in the INSPIRE project using the definitions as specified by CMS.⁶ Trained chart abstractors used a standardized chart review manual to guide abstraction of data elements. Throughout this study, we conducted a random 10 % interrater reliability assessment for all data elements included in the indicator algorithms; all kappa statistics for these data elements were >0.80.

Measure Optimization

Starting with the local VistA data, we then categorized each eCQM denominator and numerator compared to the corresponding chart review result and examined all mismatches to identify modifications to the SQL algorithms to improve eCQM performance. Iterative improvement involved generating reports of all mismatches and categorizing them as false negative or false positive. Examination of chart review data and electronic algorithm results was done to determine specific sources of error. Our group met to review all errors and discuss solutions; agreed-upon solutions were included in the next revision, and a new mismatch report was generated. In each comparison, we reviewed the number of false-negative and false-positive errors, seeking to minimize the total number of errors in the denominator and numerator separately. For example, in the STK-10 indicator we introduced SQL string searches to identify hospice discharges. Denominator false-negative results in version 2 increased from 23 to 56 (+33 errors), but denominator false-positive results fell from 101 to 51 (-50) for a net benefit of 17 fewer errors, so this text-mining strategy was carried forward into subsequent versions. We did not use natural language processing to extract free text data elements as we felt this technique would not be easily applied in subsequent real-world use cases.

Once eCQM performance had been optimized in the local VistA data, we mapped the data elements in each algorithm to

the corresponding VHA CDW data element and table. We ran the SQL queries on the cohort of stroke admissions abstracted in the INSPIRE study (N=2130, representing stroke admissions from 11 VAMCs between 2009–2012) and again compared denominators and numerators to the chart review results, categorized errors, and made iterative modifications of the algorithms as above to improve eCQM performance.

Analysis

For each eCQM denominator and numerator, we calculated the raw proportion of matched cases, the sensitivity and specificity, and the negative and positive predictive value (NPV, PPV). In the CDW sample we also characterized the overall agreement of each eCQM at the level of the stroke admission as ineligible for that indicator, eligible-passed, or eligible-failed. Since the NPV and PPV calculation is influenced by the distribution of the data, and since some of the indicators have a low proportion of eligible patients and/or a low proportion failing the indicator, the resulting PPV/NPV and agreement calculation can be highly penalized for a single mismatch.⁹ To correct for this problem, we computed both a weighted kappa and prevalence-adjusted bias-adjusted kappa (PABAK) for the overall agreement of each indicator and also computed the recommended observed and expected proportions of agreement and the prevalence and bias indices for each eCQM.^{10,11}

RESULTS

Patient characteristics of the local and CDW samples were similar except for a higher proportion of Whites and smokers, with more severe stroke (NIH Stroke Scale 6.84 vs. 5.67) and shorter length of stay in the local sample (Table 1).

The raw proportion of matched cases in each denominator and numerator ranged from 91.2 %–100 % in the local sample and from 86.4 %–99.7 % in the CDW sample (Table 2). Sensitivity was high in local (range 92.9–100 %) and CDW (86.4–99.7 %) in both numerators and denominators; specificity was high for numerators in both local and CDW (90.9–100 % and 90.8–97.0 %, respectively), but was lower for denominators (as low as 57.1 % local and 30.8 % CDW), often due to low prevalence. Positive predictive values (PPVs) ranged from 96.4 %–100 % in the local sample and from 96.9 %–100 % in the CDW sample; they were overall higher than negative predictive values (NPVs). The overall indicator agreement is shown in Table 3, with prevalence-adjusted bias-adjusted kappa values ranging from 0.73–0.95.

The range of passing rates, mean passing rates, and difference between eCQM and chart passing rates are shown in Table 4. In general, the ranges in the difference in eCQM and chart review passing rates for each indicator were of relatively small magnitude and similar direction across facilities. The only indicator with a statistically significant difference in

Table 1 Patient Characteristics

	Local cohort (N=98)	CDW cohort (N=2130)	p
Age (mean, SD, range)	66.3 (10.6) 42–93	67.1 (11.1) 26–98	0.59
% White	68.4	60.1	0.05
% Male	99.0	96.5	0.14
NIH Stroke Scale (mean, median, range)	6.8 (5) 0–36	5.6 (4) 0–38	0.01
Length of stay [mean (days), SD, range]	6.2 (5.4) 2–33	7.6 (7.1) 1–70	<0.001
% Smokers	54.1	40.4	0.01
% Hypertension	84.7	88.3	0.27
% Diabetes	50.0	42.8	0.18
% Dyslipidemia	69.4	69.5	1.00
% Prior stroke	35.7	31.8	0.44
% Prior TIA	7.1	10.1	0.49

passing rate was VTE prophylaxis, which was 10.6 % lower (worse performance) in the eCQM.

The distribution of error rates in the eCQMs is shown in Figure 1. This illustrates that each eCQM has a unique pattern of errors: almost 90 % of the errors in the VTE prophylaxis eCQM result from numerator false negatives, while more than 60 % of the errors in the NIHSS eCQM result from numerator false positives. Both of these indicators have the majority of errors resulting from numerator mismatches (error in determining passing), while the AT by HD2 and the Consider for Rehab eCQMs have the majority of errors resulting from denominator mismatches (error in determining eligibility).

The specific types of eCQM errors considering all eCQM results combined are shown in Table 5. The two largest categories of error are devices not identified (58.0 % of all numerator false-negative results) and inaccurate hospice discharge (51.7 % of all denominator false-negative results). Medication errors at discharge had multiple causes, including counting prescriptions that were either prior to or following the admission (counted as “incorrect VHA prescription”) or lack of clear documentation of discharge medications; together these accounted for 27 % of all numerator false positives.

DISCUSSION

This study demonstrates that eCQMs can be constructed with nearly 90 % or greater matching of denominator and numerator status in patients admitted with acute ischemic stroke. PPVs for these measures are generally high, suggesting that eligibility assessments are largely accurate and a passing result is likely correct. We specifically demonstrated that stroke MU indicators can be relatively accurately generated from existing EHR systems, in this case the VHA EHR. Our data also demonstrate that a relatively small number of error types is responsible for a large number of the observed mismatches, suggesting specific areas on which EHR developers and informaticians could focus to make improvements in eCQM

Table 2 eCQM Validity Results

		Local Vista				CDW			
		N	Sensitivity (% match)	Specificity	PPV/ NPV	N	Sensitivity (% match)	Specificity	PPV/ NPV
STK-1: VTE prophylaxis	Den	98	93.9 %	N/A	100.0 %	2130	99.3 %	41.2 %	99.5 %
	Num	98	99.0 %	100.0 %	N/A*	2113	86.4 %	97.0 %	33.3 %
STK-5: AT by HD2	Den	91	92.9 %	83.3 %	100.0 %	2130	98.7 %	30.9 %	99.5 %
	Num	85	100.0 %	100.0 %	90.9 %	2036	98.4 %	90.8 %	50.6 %
STK-10: consider for rehab	Den	98	96.7 %	62.5 %	98.8 %	2130	97.5 %	75.3 %	96.9 %
	Num	90	100.0 %	90.9 %	45.5 %	1948	99.3 %	94.2 %	52.7 %
STK-2: AT at discharge	Den	91	96.4 %	57.1 %	100.0 %	2130	97.6 %	76.5 %	99.1 %
	Num	84	92.8 %	100.0 %	100.0 %	1947	95.7 %	30.8 %	85.2 %
NIHSS	Den	98	93.9 %	N/A	96.7 %	2130	99.7 %	100.0 %	97.7 %
	Num	98	91.2 %	100.0 %	62.5 %	2126	98.7 %	97.1 %	73.7 %

Den=denominator; Num=numerator

accuracy for MU measures by standardizing a relatively small number of EHR data elements.

Our agreement results appear substantially better than those of previous studies outside of the VHA,^{12,13} including Persell et al.,¹⁴ who found misclassification ranging from 15–81 % for those that failed quality measures for CAD—their numbers improved when they added free text information. Two studies in the VHA did find relatively high concordance between eCQMs and chart review, similar to our results, but these looked specifically at different sets of data (in a discharge summary and the outpatient setting, respectively) rather than constructing inpatient indicators.^{15,16} Some studies report that eCQMs involving prescriptions with few contraindications seem to be captured well, with the exception of those with many documented contraindications in the text (such as warfarin for atrial fibrillation);¹⁷ however, we found that documentation of medications at the time of hospital discharge has a number of error sources. Discharge medication indicators that can be met by prescription of aspirin are complex because

it is often purchased outside of the VHA and is not always recorded in the EHR as a non-VHA medication. Devices, especially those not routinely ordered, have also been shown to be consistently difficult to capture in eCQMs, as we observed in our study.¹⁴

Although studies have shown that workflow and documentation habits affect EHR-derived quality measures,¹⁸ compared to most other studies,¹⁴ our project more fully characterized the source and types of errors in eCQMs. This detailed information can be useful to systems seeking to implement these Meaningful Use measures or to improve use of the eCQMs overall. For example, many of the errors in assessing eligibility of several eCQMs come from unstandardized discharge disposition: inaccurate hospice status and other discharge disposition errors together account for 56 % of denominator false negatives and 29 % of denominator false positives. This suggests that improving the standardization of the EHR discharge disposition assignment could improve the accuracy of multiple eCQMs. Identifying EHR changes that increase

Table 3 Overall Agreement Between Chart and eCQM Data

eCQM—chart agreement (ineligible, eligible-pass, or eligible-fail)							
	Po*	Pe†	PI‡	BI§	PABAK	K¶	95 % CI K
STK-1: VTE prophylaxis	0.87	0.68	−0.12	−0.87	0.73	0.57	0.53–0.62
STK-5: AT by HD2	0.94	0.76	−0.07	−0.88	0.87	0.59	0.36–0.82
STK-10: consider for rehab	0.94	0.68	−0.04	−0.79	0.89	0.81	0.77–0.84
STK-2: AT at discharge	0.91	0.78	0.05	−0.88	0.82	0.56	0.50–0.61
NIHSS	0.98	0.50	−0.50	−0.49	0.95	0.95	0.93–0.96

*Po=observed proportion of agreement

†Pe=expected proportion of agreement

‡PI=prevalence index

§BI=bias index

||PABAK=prevalence-adjusted bias-adjusted kappa

¶K=unweighted kappa

Table 4 Facility Passing Rates

	eCQM		Chart		Mean±SD Passing rate difference (Range of mean difference)	p
	Passing rate (range)	Mean±SD	Passing rate (range)	Mean±SD		
STK-1: VTE prophylaxis	75.8 % (45.1–91.3)	76.7±12.7	87.4 % (73.1–94.1)	87.3±6.9	–10.6±9.1 (–28.0, –2.0)	0.03
STK-5: AT by HD2	88.3 % (80.7–94.3)	88.7±4.2	90.9 % (85.3–96.2)	91.2±3.9	–2.5±3.6 (–5.5, 0.4)	0.16
STK-10: consider for rehab	88.6 % (80.4–96.0)	87.7±4.9	89.4 % (82.1–96.2)	88.3±4.9	–0.6±4.4 (–3.8, 1.7)	0.78
STK-2: AT at discharge	97.5 % (93.0–100.0)	97.4±2.1	98.0 % (96.9–100.0)	98.1±1.2	–0.7±1.5 (–12.9, 0.6)	0.35
NIHSS	48.7 % (7.4–94.5)	43.3±29.6	47.7 % (6.9–93.1)	42.3±29.8	1.0±24.4 (–1.5, 3.7)	0.94

acceptance of the measures by minimizing the type of errors that clinicians most dislike could also be helpful by reducing major sources of error in denominator false positives (i.e., incorrectly labeling ineligible patients as eligible) and reducing numerator false negatives (i.e., failing to identify an appropriately completed process). Our data would suggest that improvement of these errors may be accomplished by (1) documentation of contraindications, (2) standardized discharge disposition categories, (3) standardized device orders, and (4) more complete use of Bar Coded Medication Administration (BCMA) to document ED medication orders and medications active at discharge.

Some errors that we observed in this study are likely to be less problematic today as the VHA EHR has evolved; for example, at the time of this study there was no structured BCMA category for medication refusal. This now is part of that system, and we suspect that these errors would be reduced in a current sample of stroke admissions. Also, BCMA is

increasingly being implemented for outpatient medication administration, including VHA Emergency Departments. Finally, the documentation of non-VHA medications may have improved over time, although there remains difficulty clearly documenting medications at the time of discharge, especially when not provided by the VHA due to either lower cost non-prescription medications (e.g. aspirin) or medications prescribed by another system.

The one indicator that most extensively employed SQL text string searching performed extremely well; this may be due to the relatively unique name of the stroke severity scale (the NIH Stroke Scale) and the few and common abbreviations for this scale. The errors typically encountered in this eCQM did not represent cases of finding text that referenced something else, but rather finding text references to the NIHSS that did not indicate completion (e.g., “Couldn’t do NIHSS because patient was uncooperative”). We also used text string searches for “hospice” in the Discharge Summary, which did improve

Figure 1 eCQM error rates. Den=denominator; Num=numerator; FN=false negative, FP=false positive; N refers to the total number of errors in each eCQM

Table 5 Common eCQM Error Categories

	Denominator		Numerator	
	% False (-)*	% False (+)*	% False (-)*	% False (+)*
Device not identified	0	0	58.0	0
Inaccurate hospice D/C	51.7	8.2	0	0
Missed contraindication	0	39.2	0.6	0
Text mining error (NIHSS)	0	0	1.5	32.3
Comfort care error	11.2	16.4	0	0
Discharge disposition discrepancy	4.2	21.1	0.3	0
Med ordered not given	8.4	0	0	9.4
Missed order (BCMA not used)	1.4	9.4	9.6	0
Elective carotid intervention misidentification	19.6	0	0	0
Inaccurate consult status	0	0	3.5	12.5
D/C med documentation	0	0	10.2	13.5
Incorrect VA prescription	0	0	6.1	13.5
BCMA error	0	0	1.7	8.3
No VA prescription	0	0	7.0	0
Other	3.5	5.8	0.9	2.1

*Proportion of total number of errors in each error type accounted for by each error category

indicator performance for eligibility assessment overall (reduced denominator false positives), but also produced some false-negative results when hospice was discussed but not actually provided.

Accurate quality measurement is critical for improving stroke care and outcomes. Stroke care in the VHA is not ideal,¹⁹ and measurement is the first step toward identifying deficiencies and addressing them; however, manual chart measurement can be onerous and time-consuming, leading to delays in recognition of suboptimal care. Accurate, automated eCQMs such as those developed in this study will not only meet Meaningful Use measure requirements, but hopefully will provide early feedback to frontline providers and administrators for institution of quality improvement in stroke. This will only be possible if certain data elements are standardized to ensure the accuracy of eCQMs.

Some limitations of our study may slightly limit its generalizability to other quality measures and systems. First, the quality measures we examined did not require interfaces from outside devices such as those used in laboratory reporting, and such indicators have been reported to be problematic. Second, fixing some errors in the CDW algorithm created additional errors, which limited how much the algorithm could be optimized (e.g., text searching for “hospice” generated both false negatives and positives); however, many of these errors could be corrected with standardized data. Lastly, data collected by other EHR systems different from the VHA EHR might require a different set of algorithms,²⁰ but many of the required data elements and identified errors are likely to be similar. We did not use natural language processing (NLP) in our project, although this method may be useful to increase the accuracy of identifying highly text-based information, has a proposed

methodology to formalize NLP for clinical indicators,¹⁸ and has been used successfully in VHA projects.¹³ Although we did not include all eight stroke MU measures in this report, we are currently finalizing two additional MU indicators (anticoagulation for atrial fibrillation and statin medication at discharge), and we expect that our results will generalize to these remaining MU indicators since we need no additional sources of data to construct these final indicators. We did not attempt to construct eCQMs for the remaining two MU indicators (thrombolysis for eligible patients and stroke education) since key data elements including time of stroke symptom onset and patient-specific risk factor education documentation are not part of current VHA EHR data and would thus require substantial NLP.

CONCLUSIONS

MU indicators for stroke can be measured relatively accurately in a centralized EHR, and the relatively few sources of error could be addressed with the data standardization germane to any data system. To improve stroke MU measure accuracy, EHRs should include standardized data elements for devices, discharge disposition including hospice and Comfort Care status, recording of contraindications, and medications given in the emergency department. Future research should examine whether EHR-based indicators are cost-effective and should focus on linking these eCQMs to patient care in near real time to support care decisions and increased care quality.

Corresponding Author: Michael S. Phipps, MD, MHS; Department of Neurology/University of Maryland School of Medicine, 110 S. Paca St, 3rd floor, Baltimore, MD 21201, USA (e-mail: mhipps@som.umaryland.edu).

Compliance with ethical standards:

Conflicts of interest: Michael S. Phipps reports being on the Clinical Advisory Board for Castlight Health, a direct-to-consumer healthcare information company. Jeff Fahner reports no conflicts. Danielle Sager reports no conflicts. Jessica Coffing reports no conflicts. Bailey Maryfield reports no conflicts. Linda S. Williams, MD, reports no conflicts.

REFERENCES

1. **Leifer D, Bravata DM, Connors JJ 3rd, et al.** Metrics for measuring quality of care in comprehensive stroke centers: Detailed follow-up to brain attack coalition comprehensive stroke center recommendations: A statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42:849–877.
2. **Schwamm LH, Fonarow GC, Reeves MJ, et al.** Get with the guidelines-stroke is associated with sustained improvement in care for patients hospitalized with acute stroke or transient ischemic attack. *Circulation*. 2009;119:107–115.
3. Medicare hospital compare quality of care. Available at: <http://www.Medicare.gov/hospitalcompare/search.html>. 2015; Accessed November 2015

4. **Blumenthal D, Tavenner M.** The "meaningful use" regulation for electronic health records. *N Engl J Med.* 2010;363(6):501–504.
5. Va external peer review program. Available at: http://www.Va.Gov/vhapublications/viewpublication.Asp?Pub_id=1708. 2008; Accessed November 2015
6. **Williams L, Daggett V, Slaven JE, et al.** A cluster-randomised quality improvement study to improve two inpatient stroke quality indicators. *BMJ Qual Saf.* 2015
7. **Fihn SD, Francis J, Clancy C, et al.** Insights from advanced analytics at the veterans health administration. *Health Aff (Millwood).* 2014;33:1203–1211.
8. Commission TJ. Specifications manual for national hospital inpatient quality measures. Available at: http://www.Jointcommission.Org/specifications_manual_for_national_hospital_inpatient_quality_measures.AspX. 2015; Accessed November 2015
9. **Feinstein AR, Cicchetti DV.** High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol.* 1990;43:543–549.
10. **Byrt T, Bishop J, Carlin JB.** Bias, prevalence and kappa. *J Clin Epidemiol.* 1993;46:423–429.
11. **Viera AJ, Garrett JM.** Understanding interobserver agreement: The kappa statistic. *Fam Med.* 2005;37:360–363.
12. **Kerr EA, Smith DM, Hogan MM, et al.** Comparing clinical automated, medical record, and hybrid data sources for diabetes quality measures. *Jt Comm J Qual Improv.* 2002;28:555–565.
13. **Barkhuysen P, de Grauw W, Akkermans R, Donkers J, Schers H, Biermans M.** Is the quality of data in an electronic medical record sufficient for assessing the quality of primary care? *J Am Med Inform Assoc.* 2014;21:692–698.
14. **Persell SD, Wright JM, Thompson JA, Kmetik KS, Baker DW.** Assessing the validity of national quality measures for coronary artery disease using an electronic health record. *Arch Intern Med.* 314930544. 2006;166:2272–2277
15. **Garvin JH, Elkin PL, Shen S, et al.** Automated quality measurement in department of the veterans affairs discharge instructions for patients with congestive heart failure. *Journal for Healthcare Quality: Official Publication of the National Association for Healthcare Quality.* 2013;35:16–24.
16. **Goulet JL, Erdos J, Kancir S, et al.** Measuring performance directly using the veterans health administration electronic medical record: A comparison with external peer review. *Med Care.* 2007;45:73–79.
17. **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* 371475968. 2007;146:270–277.
18. **Parsons A, McCullough C, Wang J, Shih S.** Validity of electronic health record-derived quality measurement for performance monitoring. *J Am Med Inform Assoc.* 2012;19(4):604–609.
19. **Arling G, Reeves M, Ross J, et al.** Estimating and reporting on the quality of inpatient stroke care by Veterans Health Administration medical centers. *Circ Cardiovasc Qual Outcomes.* 2012;5:44–51.
20. **Gardner W, Morton S, Tinoco A, Scholle SH, Canan BD, Kelleher KJ.** Is it feasible to use electronic health records for quality measurement of adolescent care? *J Healthc Qual.* 2015. doi:10.1097/01.JHQ.0000462675.17265.db.