

***Grant Final Report***

**Grant ID: R18HS017067**

---

**Developing and Using Valid Clinical Quality Metrics  
for HIT with HIE**

**Inclusive Dates: 10/01/07 – 03/31/11**

**Principal Investigator:**

Rainu Kaushal, MD, MPH<sup>\*,†,‡,§,\*\*</sup>

**Team Members:**

Lisa M. Kern, MD, MPH<sup>†,‡,\*\*</sup>

Rina Dhopeswarkar, MPH<sup>\*,\*\*</sup>

Yolanda Barrón, MS<sup>†,\*\*</sup>

Sameer Malhotra, MD, MA<sup>‡,††</sup>

Jill Quaresimo, RN, JD<sup>‡‡</sup>

\* Department of Pediatrics, Weill Cornell Medical College, New York, NY

† Department of Public Health, Weill Cornell Medical College, New York, NY

‡ Department of Medicine, Weill Cornell Medical College, New York, NY

§ New York-Presbyterian Hospital, New York, NY

\*\* Health Information Technology Evaluation Collaborative, New York, NY

†† Weill Cornell Physicians Organization, New York, NY

‡‡ Taconic Independent Practice Association, Fishkill, NY

**Performing Organization:**

Weill Cornell Medical College

**Project Officer:**

Steve Bernstein

**Submitted to:**

**The Agency for Healthcare Research and Quality (AHRQ)**

**U.S. Department of Health and Human Services**

**540 Gaither Road**

**Rockville, MD 20850**

**[www.ahrq.gov](http://www.ahrq.gov)**

## Abstract

**Purpose:** To develop, validate, reliability test and use quality metrics that are electronically retrievable and appropriate for changes in ambulatory quality that health information technology, namely electronic health records (EHRs), with health information exchange (HIE) may contribute.

**Scope:** The quality metric set was developed and validated with a 36-member national expert panel and then tested and used in a federally qualified health center in New York.

**Methods:** We conducted quantitative rating of existing, relevant ambulatory quality metrics and developed metrics de novo to capture additional expected effects of EHRs with HIE. The national expert panel validated the final metric set. We determined the reliability of electronic metric retrieval on a random sample of patients whose providers used a commercial EHR and utilized the metrics to measure changes in quality over time.

**Results:** The final metric set contains 18 existing and 14 de novo metrics. We were able to assess the reliability of electronic reporting for 11 of the existing metrics (all of which are included in meaningful use) and found that the overall reliability of electronic reporting was high, with some important variation by metric. We demonstrated that provider quality of care generally improved over time with sustained use of an EHR. Overall, this work created a new set of quality metrics appropriate for EHRs and HIE, provided evidence regarding the reliability of electronic reporting of quality metrics, and suggested that quality of care may continue to improve with the use of an EHR.

**Key Words:** health information technology, electronic health records, ambulatory care, quality of care, meaningful use

<p>The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services of a particular drug, device, test, treatment, or other clinical service.</p>
--

# Final Report

## Purpose

The purpose of this project was to develop, validate, test and use a set of quality metrics that can be retrieved electronically and is appropriate for the types of improvements in quality that health information technology (HIT), namely electronic health records (EHRs), with health information exchange (HIE) may contribute in an ambulatory care setting.

The four specific aims of the project were:

- 1) To develop a modified set of quality metrics that can be retrieved electronically and is sensitive to the types of improvements in quality that HIT and HIE may contribute in an ambulatory care setting,
- 2) To validate the modified metric set with a national expert panel,
- 3) To test the reliability of electronic retrieval of the modified quality metric set, and
- 4) To use the modified quality metric set to evaluate the long-term effects of using HIT with HIE on improving health care quality.

## Scope

### Background and Context

**Overview.** Strong national policy forces are promoting the use of EHRs as an essential tool for achieving expected improvements in the quality, safety and efficiency of healthcare. Health information exchange (HIE), which enables sharing of clinical data among providers of care, is expected to significantly enhance EHRs and further improve healthcare. A number of standard metric sets exist to evaluate the clinical quality of healthcare delivered in an ambulatory care setting. However, the existing sets have not been specifically developed assuming the use of EHRs with HIE, nor do they assume that measures will be retrievable from the rich, clinical data contained within EHRs. The goal of this proposal was to develop, validate, test and use a set of electronically retrievable metrics for appropriately evaluating the impact of EHRs with HIE on the quality of clinical care delivered in the ambulatory setting.

**The Meaningful Use of Electronic Health Records.** The federal government has launched an ambitious and unprecedented program to promote the “meaningful use” of electronic health records (EHRs), which are the predominant type of HIT.<sup>1</sup> Through this program, the federal government is offering up to \$27 billion in incentives for meaningful use of EHRs with demonstrable improvements in the quality, safety and efficiency of care.<sup>2</sup> The assessment of

improvements will occur through meaningful use measures. The first set of these measures was introduced in 2010 and the next set is likely to be released in 2012. Over time, providers are expected to report their meaningful use measures to the federal government electronically, directly out of EHRs.<sup>2</sup>

Although the metric set from this grant was developed and validated prior to the establishment of the meaningful use program, the goals of this grant are fully aligned with the meaningful use program. In fact, the chair of the expert panel that approved the development and validation of our metric set was Dr. David Blumenthal, who subsequently became the National Coordinator for Health Information Technology at the U.S. Department of Health and Human Services with direct oversight of the meaningful use program.

An important component of the meaningful use program is health information exchange, which is the electronic exchange of clinical data across providers who are caring for the same patient.<sup>3</sup> Health information exchange could occur as directed communication from one provider to another, a free-standing website for providers to access community-wide data on their patients, as electronic feeds of clinical data directly into EHRs, or any of several other models of exchange. Regardless of the technical architecture, HIE is intended to improve coordination of care, which is currently very challenging to achieve in the fragmented American health care system.<sup>4</sup> Improving coordination of care is expected to increase quality and efficiency. However, the actual effects of EHRs with HIE on healthcare are not yet completely known and measurement strategies for these effects have not been clearly established.<sup>5</sup> Different models of HIE have also not been compared with each other, and the optimal strategy for HIE is unknown. Without accurate metrics that are appropriate for measuring the changes in healthcare brought about by these technologies, we will be unable to determine whether EHRs with HIE actually improve quality of care or not.

**The State of Ambulatory Quality Measurement.** There are two major limitations to existing quality metrics for measuring the effects of EHRs with HIE.

First, existing metrics were largely designed to measure quality of care in ambulatory practices as if those practices were paper-based and isolated from other providers and other care settings. As a result, they may not appropriately capture the changes in health care that may arise from the use of EHRs with HIE that supports the electronic clinical data transfer from external sources.

Second, existing quality metrics were derived with the expectation that the data would be collected from manual chart review, which is laborious and expensive, or from claims data, which lack clinical detail. Prior to meaningful use, no existing metric set had been designed for electronic reporting from clinical sources, such as EHRs, which could ultimately be less expensive than manual review and more clinically nuanced than claims.<sup>6</sup> The need to report metrics out of EHRs creates the challenge of adapting metrics appropriately for use in an EHR environment.

**The Future of Ambulatory Quality Measurement.** Even if metrics are identified for capturing the potential effects on quality of EHRs with health information exchange, it is not yet clear that automated electronic reporting of them from EHRs provide a reliable representation of the health care delivered. Specifications for numerators and denominators that were developed for claims or manual sources may not be applicable or easily adapted to electronic health record data. Electronically reported measures could either underestimate or overestimate quality,

compared to the gold standard of manual review. For example, electronically reported metrics could underestimate quality if providers' documentation of the provision of recommended care resides primarily in free-text progress notes, which are not typically amenable to automated reporting (without utilizing the advanced informatics technique of natural language processing, which is not widely available). Electronically reported metrics could also overestimate quality if, for example, they capture tests "ordered," when the specifications call for tests "completed." Therefore, once appropriate metrics are developed, they must undergo reliability testing.

Previous studies that have explored the reliability of electronic reporting found differences between electronic reporting and manual review.<sup>7-11</sup> Those studies typically addressed quality metrics for only one condition at a time<sup>7-11</sup> and were thus limited in their ability to compare the relative reliability of electronically reporting different types of metrics.

**The Importance of Measurement over Time.** Developing a valid and accurate set of electronic metrics has the potential to facilitate more nuanced measurement of quality in larger cohorts of providers over time. This is important, because use of EHRs is not static. With time, experienced users become more accustomed to an EHR with better integration into workflow. However, they may also suffer from things such as "alert fatigue" and develop a multitude of work-arounds.<sup>12</sup> Most previous assessments of the effects of EHRs on ambulatory quality have incorporated a cross-sectional design, comparing EHR users to paper users.<sup>13-18</sup> However, it is likely that, as providers become more accustomed to EHRs over time, the quality of care that they provide also changes. Electronic metrics can facilitate such longitudinal, granular measurement. This is particularly important as the second stage of meaningful use rules will be finalized over the next year and are likely to incorporate longitudinal use of EHRs.

**Summary.** Thus, EHRs with HIE are an important and evolving aspect of health care delivery. The effect of EHRs with HIE on quality are still unclear. Capturing these quality effects requires the selection of valid quality metrics and reliable reporting of those metrics from electronic sources. If the validity or reliability of electronic reporting is shown to be lacking, then any issues identified need to be addressed widely, as they are likely to arise in meaningful use reporting as well. The purpose of this project is to develop, validate, reliability test and use a set of quality metrics that can be retrieved electronically and is appropriate to the types of changes in quality that EHRs with HIE may contribute in an ambulatory care setting.

**Alignment with the Goals of the Agency for Healthcare Research and Quality.** This project addressed the overall purpose of the Ambulatory Safety and Quality (ASQ) program, which is to improve the quality and safety of health care. This project also addressed the specific goals of the Enabling Quality Measurement (EQM) arm of the program: to develop safety and quality measures in ambulatory care settings, automate quality measurement, demonstrate the ability of electronic data systems (such as EHRs) to expand potential safety and quality measures, and demonstrate improved ability to export measures for reporting and improvement. This project was also responsive to AHRQ's interest in the role of health information exchange in quality measurement.

Furthermore, this project addressed several AHRQ areas of interest for EQM grants: identify the data elements necessary for high-prevalence priority conditions, demonstrate the value and accuracy of quality measurement, demonstrate the ability of interoperable electronic systems or health information exchanges to provide data for patient safety and quality measures for episodes

of care across settings, develop the next generation electronic data systems that include the capacity for export of data, measure the efficiency of the routine measurement of ambulatory care quality and safety, and demonstrate health information exchange as a means of supplying data for and public reporting of patient safety and quality measures and emerging transparency initiatives.

## **Settings and Participants**

The first two aims of this project were addressed from a national standpoint. We conducted a literature review, which drew on the recommendations of national professional and quality measurement organizations. We also convened a panel of experts from across the country, who contributed expertise from a variety of disciplines, including health information exchange, quality measurement, health care policy and health care economics.

The first two aims also took place in the context of an ambitious initiative by New York State, which has been ongoing since 2005 (predating meaningful use) and which is investing more than \$800 million in HIT and HIE.<sup>19</sup> New York State's approach includes investment in EHRs, community-based alliances that come together for the purpose of HIE, and the Patient-Centered Medical Home.<sup>20, 21</sup> Several of the investigators on this grant are also directors (RK, LMK) and members (YB, RVD) of an independent academic collaborative (the Health Information Technology Evaluation Collaborative, or HITEC) established to evaluate the impact of New York State's initiative.<sup>22</sup> The relevance to this project is that we were able to include in our expert panel leaders from New York State communities, government and public-private partnerships, who helped ensure that our work was grounded in the realities of EHR and HIE implementation and use.

The second two aims of this project were conducted in the ambulatory setting, specifically in a federally qualified health center (FQHC) in New York State. This network serves 75,000 patients, who make 225,000 visits per year. Approximately 25% of the patients are black or Hispanic, and 75% are white. Approximately half of the patients earn incomes below the federal poverty level; more than one-third receive Medicaid and approximately one-fifth are uninsured. This setting allowed the project to capture one of the required ASQ outcome measures: the impact of HIT in a low-resourced safety net setting, where health IT diffusion is typically low.

## **Incidence, Prevalence and Alignment with Institute of Medicine Priority Areas**

This project measured quality of care for adult patients in the ambulatory setting. This project measured whether patients were receiving appropriate care for prevention, treatment and management of Institute of Medicine (IOM) priority areas (including asthma, diabetes, evidence-based cancer screening, immunization, ischemic heart disease, medication management, major depression, and severe and persistent mental illness), which is a required outcome measure for ASQ grants. This project also addressed the IOM priority area of care coordination, because it sought to identify those quality metrics that could capture the effects of HIT with HIE, a key intervention for the potential improvement of care coordination. The IOM priority areas include many of the most common illnesses and leading causes of death in the U.S.<sup>23</sup>

# Methods and Results (Aims 1 and 2)

## Study Design

Aims 1 and 2 included a literature review and national expert panel, which took place prior to the establishment of the meaningful use program.

## Data Sources/Collection

Our methodology consisted of eight steps: 1) a literature review for existing ambulatory care quality metric sets; 2) application of exclusion criteria to individual metrics contained in the metric sets; 3) articulation of assumptions, a conceptual model, and domains for rating; 4) a first round of quantitative rating; 5) validation of the process by a national expert panel; 6) a second round of quantitative rating by national experts; 7) development of metrics de novo; and 8) validation of the final metrics by national experts. More detail about these steps follows.

**Step 1.** Through the literature review, we included metric sets that were endorsed by a national quality, professional, or research organization.

**Step 2.** We then excluded individual metrics that: 1) were not relevant to ambulatory care; 2) were not relevant to adult primary care; 3) consisted of provider, practice or health plan characteristics; or 4) captured patient or provider satisfaction.

**Step 3.** We asked raters to assume the perspective of a board-certified, community-based primary care physician who has been in practice for 10 years, has a relatively stable panel of patients, and has an interoperable EHR that is currently only linked to generalist partners. We developed a conceptual model, which assumed that some relevant clinical data resided outside the physician's practice and that health information exchange could deliver those data electronically at or before the point of care. We selected two domains for a first round of quantitative rating: sensitivity to the potential effects of EHRs plus HIE and suitability for electronic reporting.

**Step 4.** Seven raters participated in the first round of rating. For each rater, we added the scores for the two domains for each metric. We then averaged the summary score across raters.

**Step 5.** We convened a 36-member national expert panel both in person and for follow-up conference calls. The panelists contributed expertise from a variety of disciplines, including health information exchange, quality measurement, health care policy and health care economics. We included representatives from community-based HIT initiatives funded by New York State, the New York State Department of Health, and the New York eHealth Collaborative. The panel was chaired by Dr. David Blumenthal, who was subsequently appointed by President Barack Obama as the National Coordinator for Health Information Technology in the U.S. Department of Health and Human Services.

**Step 6.** Sixteen expert panelists and two additional raters rated high-scoring metrics from the first round of rating along 5 domains: feasibility of delivering data electronically to the physician at the point of care, potential impact on medical decision making, clinical importance, feasibility of reporting data electronically, and a global rating.

**Step 7.** We developed metrics de novo, drawing on extensive qualitative and quantitative data obtained previously by our research group from communities across New York State that were in the process of implementing EHRs and HIEs. We circulated to communities a comprehensive list of more than 100 HIT functionalities and elements of HIE for the ambulatory setting. Communities responded by indicating which functionalities they intended to implement, along with expected clinical and economic effects. We reviewed these responses, as well as suggestions from our expert panel, to identify areas for de novo metric development. We drafted a set of metrics and presented it to the national expert panel as a whole and to 10 members individually. We iteratively refined the metrics with the experts to reach a final set.

**Step 8.** We validated this process, the top-scoring existing metrics and the de novo metrics with the national expert panel.

## **Interventions**

Aims 1 and 2 did not include any interventions.

## **Measures**

For the first round of quantitative rating, scores for individual metrics could range from 0 to 12 (that is, 0 to 6, summed across two raters). On the basis, in part, of the target number of metrics desired for the second round of rating, we considered metrics with summary scores  $\geq 9$  to be high.

For the second round of quantitative rating, scores for individual metrics could range from 0 to 6, having been averaged across raters. Choosing scores above the scale's midpoint, we considered an average score of  $\geq 4$  to be high.

Thus, the outcome measures for Aims 1 and 2 were high-scoring metrics, as described by these scoring systems, along with the metrics developed de novo.

## **Limitations**

The de novo metrics do not yet have detailed specifications, as this was beyond the scope of this project. Specifications will need to be developed in order to allow systematic and broad application of these metrics. Finally, the metrics have not yet been proven to capture the effects of EHRs with HIE; that should be the subject of future research.

## **Principal Findings and Outcomes**

We identified 17 metric sets for measuring ambulatory care quality, including metric sets by the National Quality Forum (NQF), the National Committee on Quality Assurance (NCQA), the

Doctor's Office Quality Information Technology (DOQIT) measures, the Physician Quality Reporting Initiative measures (PQRI) and others. The full list of metric sets can be found in our article in the Joint Commission Journal on Quality and Patient Safety.<sup>24</sup>

These metric sets contained a total of 1,064 individual metrics. Of these, we excluded 122 duplicates; 84 metrics not relevant to the ambulatory care setting; 136 not relevant to adult primary care; 189 consisting of provider, practice or health plan characteristics; and 23 on patient or provider satisfaction.

The remaining 510 metrics underwent the first round of ratings. The average summary score was 4.9 out of 6.0, with a standard deviation of 2.8. Fifty-nine metrics received summary scores of 9.0 or more and were thus selected for the second round of rating.

The average global score in the second round of rating was 3.8 out of 6.0, with a standard deviation of 0.6. Eighteen metrics had an average global score of 4.0 or more. One of these metrics, which dealt with medication reconciliation, was excluded by the national experts, due to concerns about feasibility. Another metric, on follow-up after discharge from a hospitalization for mental health, was added to the final set, due to the interest of the expert panel in including mental health measures.

The final 18 metrics are shown in Table 1 below.

**Table 1. Top-scoring existing metrics for measuring with electronic reporting the effect on quality of electronic health records with health information exchange (N = 18)**

	Metric Description	Original Metric Set*
1	The percentage of patients 18-56 years of age who were identified as having persistent asthma and who were appropriately prescribed medication during the measurement year. <sup>†§</sup>	NQF
2	Percentage of patients hospitalized with AMI (acute myocardial infarction) who received persistent beta-blocker treatment (6 months after discharge).	AQA
3	Patients with ischemic vascular disease who have documentation of use of aspirin or another antithrombotic during the 12-month measurement period. <sup>§</sup>	NQF
4	Patients with ischemic vascular disease whose most recent LDL-C had a result of less than 100mg/dL. <sup>§</sup>	NQF
5	Percentage of patients with HF who also have paroxysmal or chronic atrial fibrillation who were prescribed warfarin therapy.	DOQIT
6	Percentage of patients 18-75 years of age with diabetes whose most recent HbA1c level during the measurement year is >9.0%. <sup>§</sup>	NQF
7	Percentage of patients 18-75 years of age with diabetes who had one or more HbA1c test(s) during the measurement year. <sup>§</sup>	NQF
8	Percentage of diabetic patients who had at least one HbA1C measured in the reporting period below 7%. <sup>§</sup>	TCNY
9	2 part measure: Percentage of patients 18-75 years of age with diabetes whose most recent LDL-C level during the measurement year is <130 mg/dL; Percentage of patients 18-75 years of age with diabetes whose most recent LDL-C level during the measurement year is <100 mg/dL. <sup>§</sup>	NQF
10	Percentage of patients having documentation of current medication list in outpatient record.	NQF

\*NQF = National Quality Forum, AQA = Ambulatory Quality Alliance, DOQIT = Doctor's Office Quality Information Technology, NCQA = National Committee on Quality Assurance, PQRI = Physician Quality Reporting Initiative, TCNY = Take Care New York. Please see Table 1 for references for each metric set.

†This metric was modified to include only adult patients.

‡ Due to some concerns about the way it was worded, the metric from the Center for Quality Assessment and Improvement in Mental Health (CQAIMH) set was substituted with the wording from a similar metric in the Healthcare Effectiveness and Data Information Set (HEDIS). The HEDIS version was modified to include only adult patients.

§ Included in Aims 3 and 4.

**Table 1. Top-scoring existing metrics for measuring with electronic reporting the effect on quality of electronic health records with health information exchange (N = 18) (continued)**

	Metric Description	Original Metric Set*
11	Percentage of patients having documentation of allergies and adverse reactions in patient record.	NQF
12	Percentage of patients 18 years of age and older who had a follow-up visit within 30 days after being discharged for an inpatient mental health stay (including hospitalizations for depression, schizophrenia, attention deficit disorder and personality disorders). <sup>‡</sup>	NCQA
13	Percentage of patients aged 50 years and older with fracture of the hip, spine or distal radius who had a central dual-energy X-ray absorptiometry (DXA) measurement ordered or performed or pharmacologic therapy prescribed.	PQRI
14	The percentage of women 21-64 years of age who received one or more Pap tests to screen for cervical cancer. <sup>§</sup>	NQF
15	The percentage of women 40-69 years of age who had a mammogram to screen for breast cancer. <sup>§</sup>	NQF
16	The percentage of patients 65 years and older who ever received a pneumococcal vaccination. <sup>§</sup>	NQF
17	Flu shots for adults (50-64): The percentage of patients 50-64 years who received an influenza vaccination; Flu shots for older adults: The percentage of patients 65 years and older who received an influenza vaccination. <sup>§</sup>	NQF
18	Colorectal cancer screening by colonoscopy performed (Age 50-80). <sup>§</sup>	TCNY

\*NQF = National Quality Forum, AQA = Ambulatory Quality Alliance, DOQIT = Doctor's Office Quality Information Technology, NCQA = National Committee on Quality Assurance, PQRI = Physician Quality Reporting Initiative, TCNY = Take Care New York. Please see Table 1 for references for each metric set.

†This metric was modified to include only adult patients.

‡ Due to some concerns about the way it was worded, the metric from the Center for Quality Assessment and Improvement in Mental Health (CQAIMH) set was substituted with the wording from a similar metric in the Healthcare Effectiveness and Data Information Set (HEDIS). The HEDIS version was modified to include only adult patients.

§ Included in Aims 3 and 4.

We identified 5 topics not well represented among existing metrics: test ordering, medication management, referrals, follow-up after discharge, and revisits. We developed 14 metrics to address these topics, as shown in Table 2 below.

**Table 2. Metrics developed de novo for measuring with electronic reporting the effect on quality of electronic health records with health information exchange (N = 14)**

#	Metric
1	<p><u>Repeat Blood Tests</u></p> <p>For each type of blood test below, consider: Of all the tests ordered by a provider over a six-month period, how many represent tests for which results were already completed for that patient (regardless of the ordering provider) and are less than [insert appropriate repeat interval] old at the time of the second test?</p> <p>Hemoglobin (10 days)            Creatinine (10 days)            Sodium (10 days)            Total cholesterol (6 weeks)            HDL cholesterol (6 weeks)            Thyroid stimulating hormone (6 weeks)            Liver function tests (ALT/AST) (6 weeks)            Ferritin (8 weeks)            Hemoglobin A1c (12 weeks)</p>

**Table 2. Metrics developed de novo for measuring with electronic reporting the effect on quality of electronic health records with health information exchange (N = 14) (continued)**

#	Metric
2	<u>Repeat Imaging Studies</u> Of those imaging studies (x-rays, ultrasounds, CT scans, and MRIs) ordered by a provider over a three-month period, how many represent tests for which results were already completed for that patient (regardless of the ordering provider) and are no more than 60 days old at the time of the second test?
3	<u>Repeat Cardiac Studies</u> Of those cardiac studies (all variants of stress tests and echocardiography) ordered by a provider over a three-month period, how many represent tests for which results were already completed for that patient (regardless of the ordering provider) and are no more than 90 days old at the time of the second test?
4	<u>Generic Prescribing</u> Of all medications prescribed by a given provider and filled by patients over a three-month period, how many are filled as generic?
5	<u>Formulary Prescribing</u> Of all medications prescribed by a given provider and filled by patients over a three-month period, how many are on formulary?
6	<u>Fill Data</u> Of all patient visits to a given provider over a three-month period, how many have fill data available at the point of care?
7	<u>Discharge Medication Documentation</u> Of all patients discharged from a hospital over a three-month period, how many patients have the discharge medication list documented by the primary care provider during the first outpatient visit following discharge (which could occur up to one month after the discharge)?
8	<u>Reason for Referral</u> Of all patients referred to a specialist by a primary care physician over a three-month period, how many have the primary care physician's reason for the referral sent to the specialist's office?
9	<u>Specialist Recommendations</u> Of all patients referred to and seen by a specialist for a given primary care physician, for how many patients is the specialist's recommendations sent back to the primary care physician by the time of the patient's next follow-up appointment with the primary care physician (which could occur up to three months after the referral visit)?
10	<u>Post-Discharge Hospital Follow-up</u> Of all patients who are hospitalized in a three-month period, how many are seen by their primary care physicians within 14 days of discharge?
11	<u>Hospital Discharge Summary</u> Of all patients who are hospitalized in a three-month period, how many had a discharge summary received by their primary care physician within 14 days of discharge?
12	<u>Ambulatory Care-Sensitive Conditions ED Visits*</u> Of all patients with at least one ambulatory care-sensitive condition that are in a physician's panel at a given time, how many are seen in the emergency department over the subsequent three months for that ambulatory care-sensitive condition?
13	<u>Ambulatory Care-Sensitive Conditions Hospitalizations Visits*</u> Of all patients with at least one ambulatory care-sensitive condition that are in a physician's panel at a given time, how many are hospitalized over the subsequent three months for that ambulatory care-sensitive condition?
14	<u>Readmissions</u> Of all patients who are hospitalized in a three-month period, how many are readmitted within 30 days of discharge?

\* Common ambulatory care-sensitive conditions for adults, as defined by the Agency for Healthcare Research and Quality, include: chronic obstructive pulmonary disease, adult asthma, congestive heart failure, angina without procedure, hypertension, diabetes with long-term complications, diabetes with short-term complications, lower extremity amputations among patients with diabetes, uncontrolled diabetes without complication, bacterial pneumonia and urinary tract infections requiring hospitalization

## Conclusions and Discussion

We developed a set of 18 metrics, selected from among metrics already endorsed by national organizations. We also generated a set of 14 metrics de novo, which address the portion of

quality that overlaps with utilization and that targets care coordination more explicitly than existing metrics.

The process of developing and validating these metrics raised 5 issues that are highly relevant to the current national discussion on EHRs and quality.

- 1) Data structure: Quality reporting activities are most likely to focus on those variables captured in structured fields until natural language processing becomes more sophisticated and routine. Thus, the method of data structure plays an important role in shaping what is possible to measure in an electronic environment.
- 2) EHR usability and workflow: Although structured fields can be used to generate electronic reports, they are unlikely to add value unless health care providers consistently use them. Creating easy-to-use computer interfaces, restructuring clinical workflow, and having comprehensive training, configuration and ongoing technical support are all likely to affect the reliability of electronically reported metrics.
- 3) Community integration: The reliability of electronic quality reporting for measuring the effects of HIE will depend not only on the capacity of an EHR to receive external clinical data but also on the prevalence and completeness of community-wide data exchange.
- 4) Vendor maturity and priorities: There is still a large gap between what one would like to report electronically and what is currently possible. At the time of our study, out of the many vendors we considered, none were able to electronically report all 18 metrics that our national panel endorsed. We subsequently identified an FQHC that could report out 12 of the metrics, because it had been pursuing electronic reporting on its own for quality improvement purposes. The FQHC had its own internal IT staff that built custom electronic reports and even that proved to be a challenging process.
- 5) Quality metric specification: Enabling electronic reporting is partly a programming issue but also involves developing specifications (definitions for numerators, denominators and exclusion criteria for each metric) that are applicable to EHRs.

## Significance and Implications

The development and validation of this metric set predated meaningful use and was able to inform the measures for meaningful use. We had the privilege of having David Blumenthal, MD, MPP, moderate our national expert panel in 2008, before he became the National Coordinator for Health Information Technology. As the National Coordinator, his office oversaw the establishment and development of the meaningful use program, including measure development. Our metrics were published in the Joint Commission Journal on Quality and Patient Safety in 2009,<sup>24</sup> before the release of meaningful use metrics in 2010. Our metric set was known to and helped inform the Office of the National Coordinator's (ONC) Health IT Policy Committee at the time it developed meaningful use metrics.

Of the 18 existing metrics in our final set, 15 are included in Stage 1 meaningful use in identical or similar forms (with slight variations based on the wording of the metric or the clinical cutoff used). The ones not included relate to: hemoglobin A1c testing (though that is

represented, in a sense, by other metrics that focus on A1c control), follow-up after discharge from a mental health-related hospitalization, and diagnosis of osteoporosis. None of the metrics developed de novo are reflected in Stage 1 meaningful use, in part because they are novel metrics that do not yet have accompanying specifications. In sum, the existing metrics in our study are well aligned with and support the conceptual underpinnings of meaningful use.

The next stage of meaningful use metrics are currently under development. Dr. Kaushal, the Principal Investigator, serves on the Care Coordination Tiger Team for ONC, which informs the Health IT Policy Committee regarding meaningful use. In that capacity, she has been asked to present the de novo metrics.

Quality measurement, particularly measurement that is enabled by EHRs and is sensitive to the potential effects of EHRs with health information exchange, has never been more important. Adoption of EHRs is increasing steeply with the meaningful use program. EHRs have the potential to transform both the delivery and measurement of health care. Development and validation of appropriate metrics, such as those developed here, is critically important to determine the effects of EHR use on health care quality. Furthermore, valid quality measurement can inform iterative refinements of EHR-based clinical decision support and quality improvement efforts.

## **Methods and Results (Aims 3 and 4)**

### **Study Design**

Aim 3 was addressed with a cross-sectional study, while Aim 4 was addressed with a longitudinal cohort study.

### **Data Sources/Collection**

For Aim 3, we partnered with an FQHC that had adopted a commercially available EHR in 2002. From the modified quality metric set derived from Aims 1 and 2, we selected the 12 metrics that the FQHC was able to report electronically. Of these, 11 are included in Stage 1 meaningful use as Clinical Quality Measures. The 12<sup>th</sup>, percentage of patients who had at least one hemoglobin A1c measured, is not included as a separate meaningful use metric but is implied as part of the two hemoglobin A1c control measures (that is, percentage of patients with diabetes whose hemoglobin A1c is <7%, and percentage of patients with diabetes whose hemoglobin A1c >9%).

Our study population included all patients 18 years and older who had at least one office visit with a general internist or family practitioner at one of the FQHC sites north of New York City during calendar year 2008. We identified electronically which patients were eligible for the quality metrics and took a random sample of those. For this sample, we collected the same 12 quality metrics in two ways: automated electronic reports and manual review. The manual reviewers were two members of our research team: a physician-informaticist and a research nurse whom we trained for 5 hours each, specifically for this project. Their inter-rater reliability scores were high, with kappa scores of 0.74, 0.84, 0.89, 1.00, 1.00 and 1.00 for the metrics tested.

The inter-rater reliability statistics were discussed with the reviewers, and the project leaders were available for questions during the data collection process.

For Aim 4, we used the same metrics as in Aim 3. We started with the study population described above for 2008 and generated a similar study population for 2009. We included primary care providers (family medicine physicians, general internists, pediatricians, obstetrician-gynecologists, and nurse practitioners) who had patients in both years. We restricted the dataset further to primary care providers with at least 25 patients in their panels who were eligible for these quality metrics in both years. We then generated electronic reports for both years.

## **Interventions**

Aim 3 evaluates the EHR as an electronic documentation and reporting tool, and Aim 4 evaluates whether the EHR influences quality over time. The FQHC views the purpose of the EHR to be quality improvement. The FQHC reports that it is one of the first free-standing community health center networks in the country to implement an integrated EHR and practice management system. The EHR design is that of a commercially available product, which is certified by the Certification Commission for Health Information Technology and includes clinical decision support. It was implemented in 2002 and is used by all providers in that setting.

The EHR also includes HIE functionalities, including interoperability with laboratory, radiology and pathology results. All of the FQHC's clinical operations and information technology are fully integrated across sites and centrally managed. This adds another layer of interoperability, with access to the same clinical data across disparate geographic sites. All of the clinical departments participate in regular quality improvement projects, which are enabled by electronic reports generated from the EHR by an experienced, internal IT staff. With a web-based user network, the FQHC also pursues projects recommended by other health systems across the country that have used the same EHR product. The FQHC also has its own secure electronic patient portal, through which patients can view their medical records and communicate with providers.

## **Measures**

For Aim 3, we created two-by-two tables for each metric, comparing the electronic reports to the gold-standard manual review in order to determine the reliability of electronic reporting. The rows of these tables showed whether the patients met criteria for the metric numerator (that is, received the recommended care), according to the electronic reports. The columns showed whether the patients met criteria for the numerator according to manual review. We counted the numbers of true positives, true negatives, false negatives and false positives for each metric. We then calculated the sensitivity and specificity for each measure. Sensitivity expresses the ability of electronic reporting to detect recommended care when the recommended care was given, whereas specificity expresses the ability of electronic reporting to correctly identify patients who did not receive the recommended care. We also calculated a summary measure of sensitivity and specificity across metrics.<sup>25</sup> In addition, we calculated the proportion of patients who received recommended care, according to each data collection method (electronic reporting vs. manual review). We determined whether electronic reporting generated an overestimate, underestimate or similar estimate of quality, compared to manual review.

For Aim 4, we calculated for each provider the percentage of his/her patients who received recommended care for each metric. We then used a mixed model to determine whether quality overall improved over time, adjusting for repeated measures over time and for clustering of metrics within provider.

## Limitations

The limitations of Aims 3 and 4 include that they are conducted in a single FQHC, which may not be generalizable to other FQHCs or other health care settings. The FQHC we studied had been using an EHR for 6 years and had its own internal IT staff. While this may limit the study's generalizability, it also suggests that our estimates of the reliability of electronic reporting may be higher than those that could be generated elsewhere.

## Principal Findings and Outcomes

For Aim 3, we studied 1154 patients who were cared for by 106 providers. Approximately two-thirds of the patients were female. The average age was 55 years old. Three-fourths were white. Approximately 39% had Medicare, 32% had commercial insurance, 16% had Medicaid and 7% were uninsured.

We obtained electronic reports on 12 metrics, but ultimately excluded from the analysis the metric related to mammography screening, as we encountered technical difficulties relating to reporting that metric, in terms of distinguishing between mammograms ordered and mammograms completed. We were able to consistently capture mammograms ordered, but the metric specified mammograms completed.

For the remaining 11 metrics, we found that the overall reliability of electronic reporting was high. Electronic reporting had a sensitivity of 0.88 (95% confidence interval 0.75, 0.94), meaning that it was able to detect 88% of the patients who received recommended care. Electronic reporting had a specificity of 0.89 (0.81, 0.94), meaning that it was able to correctly identify 89% of the patients who did not receive recommended care.

Although reliability was high overall, it varied considerably across metrics. Sensitivity ranged from 46% (for appropriate asthma medication) to 98% (for having a hemoglobin A1c test done). Specificity ranged from 62% (for appropriate antithrombotic medication for patients with ischemic vascular disease and for LDL control among patients with diabetes) to 97% (for pneumococcal vaccination).

When we calculated the proportion of patients who received recommended care by each data collection method (electronic reporting and manual review), we found that electronic reporting underestimated quality for 1 metric, overestimated quality for 4 metrics, and generated similar estimates for 6 metrics. A manuscript describing Aim 3 is in preparation and will be submitted for publication shortly.

For Aim 4, we included 45 primary care providers. Two-thirds were attending physicians, and more than three-fourths were in family medicine. These providers cared for 10,400 patients who were eligible for the 11 quality metrics in 2008 and 2009. During the analysis stage, we excluded one of the 11 metrics, the metric on the percentage of patients with diabetes who had hemoglobin A1c's >9%, as the electronic report was internally inconsistent for 2009 only. Quality of care improved from 2008 to 2009 for all 10 of the remaining quality metrics, with the absolute improvement ranging from 2.2 percentage points (for asthma medication) to 33.1

percentage points (for colorectal cancer screening). The mean absolute improvement was 14.9 percentage points. Overall, quality significantly improved over time ( $p < 0.0001$ ). A manuscript describing Aim 4 is also in preparation and will be submitted shortly.

## Conclusions and Discussion

We found that the overall reliability of electronic reporting was high, but that reliability varied considerably across metrics. We encountered substantial problems with electronic reporting of one metric in 2008 and two metrics in 2009, such that they had to be excluded from analysis, at least until the technical problems can be solved.

Several patterns were observed that are worth noting. Metrics that involved medications (e.g. asthma, ischemic vascular disease) tended to have lower reliability, in contrast to metrics that involved laboratory tests (e.g. diabetes), which tended to have higher reliability. Medications are frequently recorded in a combination of structured and unstructured (free text) forms and in multiple places of the medical record, whereas laboratory test results are more consistently structured, especially in the context of HIE. Tests that are done by specialists (e.g. colorectal cancer screening) tended to have lower reliability, perhaps because they were contained in text-based reports not amenable to electronic reporting from structured fields. Some vaccines were easier to report (e.g. influenza) than others (e.g. pneumococcus), perhaps because they are given at different frequencies (yearly versus once) and may be tracked differently (in structured flowsheets or text-based fields).

These variations underscore the complexity of electronic reporting and the need for attention to detail in generating the programming that underlies electronic reports. NQF is currently developing standardized specifications for electronic reporting.<sup>26</sup> To our knowledge, their specifications have not yet been validated against manual review, and our results suggest that that would be an important step. NQF is aware of our work.

## Significance and Implications

This is the first study to our knowledge to test the reliability of electronically reporting meaningful use metrics. The reliability of electronic reporting was high overall. However, the reliability varied widely by metric and two metrics were ultimately excluded from analysis, due to careful detection of internal inconsistencies that could easily be missed by providers using electronic reporting for operational purposes. In addition, the paucity of vendors able to generate only a short while ago the needed electronic reports illustrates how far the field needs to evolve to meet the demands of meaningful use.

This is also the first study to our knowledge that demonstrates improvements in quality with sustained use of EHRs. Previous studies of this kind have been few and have generally found no effect.<sup>27, 28</sup> This is important, because it bolsters the rationale for meaningful use of EHRs and does so with electronic reports of quality measures as will be measured in this program.

Due to the meaningful use program, the entire country is moving toward electronic reporting of quality measures. If quality reporting is not valid, though, the financial incentives (and financial penalties) could be given to the wrong providers. Overall, this work newly establishes the reliability of electronic reporting of specific meaningful use metrics, and suggests that quality of care may continue to improve with the use of an EHR with HIE.

## References

1. U.S. Department of Health and Human Services. Medicare and Medicaid Programs; Electronic Health Record Incentive Program; Final Rule. 75 Federal Register 44314 (2010) (42 CFR Parts 412, 413, 422 and 495).
2. Blumenthal D, Tavenner M. The "Meaningful Use" Regulation for Electronic Health Records. *N Engl J Med* 2010;363:501-4.
3. Blumenthal D, Glaser JP. Information technology comes to medicine. *N Engl J Med* 2007;356:2527-34.
4. Pham HH, O'Malley AS, Bach PB, Saiontz-Martinez C, Schrag D. Primary care physicians' links to other physicians through Medicare patients: the scope of care coordination. *Ann Intern Med* 2009;150:236-42.
5. Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006;144:742-52.
6. Pawlson LG. The past as prologue: future directions in clinical performance measurement in ambulatory care. *Am J Manag Care* 2007;13:594-6.
7. 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;146:270-7.
8. 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;28:555-65.
9. Kmetik KS, O'Toole MF, Bossley H, et al. Exceptions to outpatient quality measures for coronary artery disease in electronic health records. *Ann Intern Med* 2011;154:227-34.
10. Persell SD, Wright JM, Thompson JA, Kmetik KS, Baker DW. Assessing the validity of national quality measures for coronary artery disease using an electronic health record. *Arch Intern Med* 2006;166:2272-7.
11. Weiner M, Stump TE, Callahan CM, Lewis JN, McDonald CJ. Pursuing integration of performance measures into electronic medical records: beta-adrenergic receptor antagonist medications. *Qual Saf Health Care* 2005;14:99-106.
12. van der Sijs H, Aarts J, van Gelder T, Berg M, Vulto A. Turning off frequently overridden drug alerts: limited opportunities for doing it safely. *J Am Med Inform Assoc* 2008;15:439-48.
13. Friedberg MW, Coltin KL, Safran DG, Dresser M, Zaslavsky AM, Schneider EC. Associations between structural capabilities of primary care practices and performance on selected quality measures. *Ann Intern Med* 2009;151:456-63.
14. Keyhani S, Hebert PL, Ross JS, Federman A, Zhu CW, Siu AL. Electronic health record components and the quality of care. *Med Care* 2008;46:1267-72.
15. Linder JA, Ma J, Bates DW, Middleton B, Stafford RS. Electronic health record use and the quality of ambulatory care in the United States. *Arch Intern Med* 2007;167:1400-5.
16. Poon EG, Wright A, Simon SR, et al. Relationship between use of electronic health record features and health care quality: results of a statewide survey. *Med Care* 2010;48:203-9.
17. Romano MJ, Stafford RS. Electronic Health Records and Clinical Decision Support Systems: Impact on National Ambulatory Care Quality. *Arch Intern Med* 2011;48:203-9.
18. Walsh MN, Yancy CW, Albert NM, et al. Electronic health records and quality of care for heart failure. *Am Heart J* 2010;159:635-42 e1.
19. New York State Department of Health. Office of Health Information Technology Transformation. 2010. (Accessed June 29, 2011, at <http://www.health.state.ny.us/technology/>)
20. Kern LM, Barron Y, Abramson EL, Patel V, Kaushal R. HEAL NY: Promoting interoperable health information technology in New York State. *Health Aff (Millwood)* 2009;28:493-504.
21. Kern LM, Wilcox AB, Shapiro J, et al. Community-based health information technology alliances: potential predictors of early sustainability. *Am J Manag Care* 2011;17:290-5.
22. Kern LM, Kaushal R. Health information technology and health information exchange in New York State: new initiatives in implementation and evaluation. *J Biomed Inform* 2007;40:S17-20.

23. Agency for Healthcare Research and Quality. Priority areas for national action: Transforming health care quality: Summary of Institute of Medicine report. 2003. (Accessed June 29, 2011, at [www.ahrq.gov/qual/iompriorities.htm](http://www.ahrq.gov/qual/iompriorities.htm).)
24. Kern LM, Dhopeswarkar R, Barron Y, Wilcox A, Pincus H, Kaushal R. Measuring the effects of health information technology on quality of care: a novel set of proposed metrics for electronic quality reporting. *Jt Comm J Qual Patient Saf* 2009;35:359-69.
25. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982-90.
26. National Quality Forum. Electronic quality measures (eMeasures). 2011. (Accessed June 28, 2011, at [http://www.qualityforum.org/Projects/e-g/eMeasures/Electronic\\_Quality\\_Measures.aspx](http://www.qualityforum.org/Projects/e-g/eMeasures/Electronic_Quality_Measures.aspx).)
27. Garrido T, Jamieson L, Zhou Y, Wiesenthal A, Liang L. Effect of electronic health records in ambulatory care: retrospective, serial, cross sectional study. *BMJ* 2005;330:581.
28. Zhou L, Soran CS, Jenter CA, et al. The relationship between electronic health record use and quality of care over time. *J Am Med Inform Assoc* 2009;16:457-64.

## List of Publications and Products

### Publications

Kern LM, Blumenthal D, Pincus H, Dhopeswarkar R, Kaushal R. Quality measures for capturing the effects of health information exchange [abstract]. *AMIA Annu Symp Proc* 2008 Nov 6:1001.

Kern LM, Dhopeswarkar R, Barrón Y, Wilcox A, Pincus H, Kaushal R. Measuring the effects of health information technology on quality of care: A novel set of proposed metrics for electronic quality reporting. *Jt Comm J Qual Patient Saf* 2009;35(7):359-69.

Kern LM, Kaushal R. Interoperable electronic health records and quality of care [expert commentary]. June 2010. National Quality Measures Clearinghouse, Agency

for Healthcare Research and Quality, Rockville, MD. <http://qualitymeasures.ahrq.gov/expert/expert-commentaries-index.aspx>.

Kern LM, Malhotra S, Barrón Y, Quresimo J, Dhopeswarkar R, Kaushal R. Are electronically-reported “Meaningful Use” clinical quality measures valid? To be submitted to the *Annals of Internal Medicine*.

Kern LM, Barrón Y, Dhopeswarkar R, Kaushal R. Improvements in ambulatory quality with sustained use of electronic health records. To be submitted to the *Am J Manag Care*.

### Related Publications

Kern LM, Barrón Y, Blair AJ III, Salkowe J, Chambers D, Callahan M, Kaushal R. Electronic result viewing and quality of care in small group practices. *J Gen Intern Med* 2008;23:405-10.

Bates DW. The approaching revolution in quality measurement [editorial on our article]. *Jt Comm J Qual Patient Saf* 2009;35(7):358.

Kern LM, Barrón Y, Dhopeswarkar R, Kaushal R. Health information exchange and quality of care. Under review at the *J Gen Intern Med*.

Kern LM, Barrón Y, Dhopeswarkar R, Kaushal R. Electronic health records and ambulatory quality of care. Submitted to *JAMA*.

## Presentations

Oral presentation, “Developing a metric set for measuring and reporting ambulatory quality of care in the setting of health information technology with health information exchange,” Annual Conference of the Agency for Healthcare Research and Quality, Bethesda, MD (September 8, 2008)

Poster, “Quality measures for capturing the effects of health information exchange,” Annual Symposium of the

American Medical Informatics Association, Washington, DC (November 6, 2008)

Lead panel presentation (along with presentations by Paul Tang, MD, MS and Jonathan Weiner, DrPH), “Quality Reporting from Health Information Technology: Challenges, Opportunities and Insights,” Annual Symposium of the American Medical Informatics Association, Washington, DC (October 2011, accepted)

# Appendixes

## Appendix A

### Relevance to AHRQ Priority Populations

This project included electronic reporting of quality metrics in a federally qualified health center that provides care to several of AHRQ's Priority Populations, including low-income groups, minority groups, women, and the elderly. The study setting and the patient population are described in more detail in the Scope section and in the Results/Aims 3 and 4 section.

## Appendix B

### Fulfillment of AHRQ Requirements for Ambulatory Safety and Quality (ASQ) Final Progress Reports for “Enabling Quality Measurement (EQM) through Health IT Goals” Grants

This project is directly responsive to the goals of AHRQ, the ASQ program overall and the EQM grants in particular. This grant demonstrates an improvement in quality over time in an ambulatory setting for low-income populations. This grant advances the field of quality measurement and, in particular, advances the field of electronic reporting of quality by: 1) demonstrating the ability of electronic data systems to generate quality measures, 2) measuring the reliability of electronic reporting, and 3) measuring changes in quality over time. This project also defines a set of measures that can be used to capture the effect of HIE, which addresses the IOM priority area of care coordination.

The work done through this grant is scalable and has contributed to the sustainability of work in quality measurement. The quality metric set has already been made scalable through the meaningful use program. The field of quality measurement will be increasing in scope as a result of this program. The next generation of projects will need to focus on the identification of additional quality metrics appropriate for EHRs and HIE, and on how to confirm in other settings that specific metrics are appropriately and accurately reported from EHRs and/or HIEs. In addition, further work is needed to develop and test quality metrics that more fully utilize the clinically rich data in EHRs and HIEs. Our experience with this grant and the results we have generated suggest that further work is needed in this important and expanding field of EHR- and HIE-based quality measurement, particularly in the context of the meaningful use program.