Grant Final Report

Grant ID: R18HS017160

Cardio-HIT Phase II

Inclusive Dates: 09/30/07 - 12/31/09

Principal Investigator:
Karen S. Kmetik, PhD[†]

Team Members:
Michael O'Toole, MD, co-PI[‡]
Heidi Bossley, MSN, MBA[†]
Rachel Connolly**
Bridget Gulotta, MSN, MBA[†]
Sherry L. Grund, RN**
Mark Hennessey, MBA**
L. Greg Pawlson, MD, MPH[§]
Joan Roberson**
Patricia A. Schwamberger, RHIA[‡]
Sarah Scholle, MPH, DrPh[§]
Gregory Wozniak, PhD[†]
Carmen A. Brutico, MD, MBA*, [††]
John Carroll*, [‡‡]
Marsha Dunlap, RN*, [‡‡]
Gary Fischer, MD*, [§§]
Patricia Gallacher*, [††]
Stasia Kahn, MD*, ***
Karen M. Murphy, RN, PhD*, [††]
Ted Pacheco, MD*, [‡‡]
Dave Pietrzyk*, [§§]
John Schaeffer, MD*, [‡‡]
Vikram Sheshadri*, ***

* Practice Site Team Members
[†] American Medical Association
[‡] Midwest Heart Specialists
[§] National Committee for Quality Assurance
** Iowa Foundation for Medical Care
[††] Physician Health Alliance
[‡‡] North Ohio Heart Center
[§§] University of Pittsburgh Medical Center
*** Fox Prairie Medical Group

Performing Organization:
American Medical Association

Project Officer:
Jeff Brady

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

Abstract

Purpose: Cardio-HIT Phase II investigated and validated the prevalence and patterns of exception reporting among physicians participating in the Cardio-HIT collaborative, a type of practice-based research network.

Scope: The Cardio-HIT Phase II collaborative was formed as a research network of five independent physician group practices to demonstrate the feasibility and value of implementing the nationally-recognized, ambulatory care physician performance measures for coronary artery disease and heart failure within existing practice site electronic health record systems. The five physician practices submitted clinical performance data to a data warehouse for coronary artery disease and heart failure and received de-identified, aggregate performance and exception reports.

Methods: Cardio-HIT Phase II was a non-experimental, observational study designed to investigate the prevalence and patterns of rates of exceptions, apparent quality failures and measure met performance reported electronically for the coronary artery disease and heart failure measures. This research was a descriptive study for generating hypotheses for future research rather than a hypothesis-testing trial.

Results: Performance rates were generally high with exception reporting generally low. Agreement rates between automatic submission and manual review were also high. Because many exceptions are not absolute, physicians may decide to "over-ride" an exception and provide the relevant aspect of care.

Key Words: exception reporting, performance measures, quality indicators, electronic health record systems

Final Report

Purpose

Objectives of Study

The objectives of this study were to assess the prevalence of exception reporting, document specific reasons for exceptions, evaluate the relative accuracy of reported exceptions, and identify the location of exception data in electronic health record systems (EHRs). Exception data were collected, categorized as medical, patient, or system reasons, and reported separately with performance rates. Many of the physician performance measures developed by the American Medical Association (AMA)-convened Physician Consortium for Performance Improvement® (PCPI) allow physicians to identify patients with exceptions (eg, drug allergy) as a measure to track variations in care.

More specifically, in the context of physician performance measurement, exceptions are an element of measure design that serve 3 critical functions in physician performance measurement: (1) they align the intent of the measures with a physician's need to exercise clinical judgment and to include consideration of patient preferences in providing care; (2) they may promote case-by-case consideration of appropriate care, and thus quality of care; and (3) they create more homogeneity in the denominator populations to enhance comparability, and thus reduce the possibility of misinterpreting results for physicians who treat severely ill or vulnerable populations. All of these elements are central to obtaining physician acceptance of quality performance data. The most commonly used alternative to allowing direct exceptions is to use case mix and risk adjustment in performance measurement. However, the complexity introduced by risk adjustment, the need for larger sample sizes and the difficulty in collecting required data elements raise serious concern—especially among clinicians—about the feasibility of case mix and risk adjustment to improve measurement.[1]

In the measures created by the PCPI, which are in widespread use in the voluntary Physician Quality Reporting Initiative (PQRI) of the Centers for Medicare & Medicaid Services (CMS), the specifications allow physicians to identify and compare patient populations in the denominator of performance measures (process measures) using exceptions as an alternative approach to risk adjustment. These exception criteria use broad categories; namely, exception for *medical, patient or system reasons*. Some examples of the specific exceptions that would be expected in each of these categories are noted in the measures (eg, patient exception because of patient refusal of treatment), but it is left to the judgment of the physician whether a patient should be excluded from the denominator of the measure. In a pay for reporting program, such as the CMS PQRI or the proposed rule on the adoption and meaningful use of electronic health record technology, physicians receive "equal credit" for patients who receive the appropriate care (e.g., patients who are prescribed a particular medication, as specified by a measure) and those with a valid exception (e.g., did not receive medication because of a documented allergy).

With the growing levels of public and private clinical performance measurement activity in the U.S., the use of exceptions has received increasing attention from a number of different stakeholders.

- Patients place their trust in their healthcare providers to deliver appropriate and safe care. Patients expect physicians to document and recall their preferences and clinical information in a manner that is conducive to appropriate decision-making. A key question is whether the system of exception reporting—and its implicit documentation requirements—aids in this decision-making and ultimately results in improved quality of care.

- Physicians are concerned about the validity of performance measures and that the measurement process yields fair evaluations of true quality. A key question is whether exception criteria and exception reports are accurate and valid, as well as what impact exception reporting has on overall performance measurements and physician acceptance of performance evaluations.

- Payers and policymakers interested in performance measurement for value-based purchasing are concerned about potential loopholes for "gaming" by physicians that would defeat the purpose of performance measurement. If exceptions are applied too generally and non-uniformly, biases in measurement could occur. Key questions for these stakeholders include the prevalence and distribution of exception reporting, how outliers can be identified, and whether outliers are indicative of gaming or some other system factor.

- Performance measure developers seek feasible, practical ways of providing performance measurement and often have to balance attaining higher accuracy against the cost and complexity of gathering data. At the same time, measures should adequately account for patient heterogeneity. A key question is whether exception criteria can be applied accurately to identify appropriate denominators (ie, a set of eligible patients) for measurement.

- Product vendors of EHRs will play a critical role in shaping performance measurement through product design. Vendors need to assure physicians that their products facilitate identification of all patients with exceptions. A key question is whether EHRs can be designed to facilitate collection of exceptions in a way physicians find useful.

To date, the researchers are aware of no systematic investigations into these key issues concerning exception reporting that emphasize the use of different EHRs at different independent practice sites. The overarching objectives of the study were to advance the science of performance measurement through quantitative study of the prevalence and patterns of exception and performance reporting, advance our empiric knowledge of the relative accuracy of exceptions, help inform the national debate on the role of exception reporting in physician performance measurement through a qualitative study of key stakeholder perspectives on exception reporting in performance measurement, and develop ways to better delineate patient populations through more detailed exception categorization.

This observational study of exception reporting used data from *Cardio-HIT Phase I*, a research collaborative of 6 EHRs-enabled independent group practices that collected data and reported on the American College of Cardiology (ACC)/American Heart Association (AHA)/PCPI measures for coronary artery disease (CAD) and heart failure (HF).

To advance the science of performance measurement, 3 specific aims were addressed:

Specific Aim 1. Develop an empirical understanding about prevalence and patterns of exception reporting among physicians using EHRs and reporting national performance measurements. Exception and performance reporting data were used from the Cardio-HIT sites to quantify prevalence and patterns of exceptions and performance for 2 measure sets: CAD and HF.

Specific Aim 2. Evaluate the feasibility and accuracy of exception reporting among physicians. To achieve this aim, (a) organizational evaluations were conducted to characterize and assess the ability of EHRs-enabled practices to capture data required for exception reporting, and to assess variation in this process; and (b) evaluate the accuracy and validity of automated exception reports and identify key sources of measurement error.

Specific Aim 3. Analyze and then address stakeholder perspectives concerning exception reporting in physician performance measurement to develop refined principles and methods regarding the use of exception reporting in performance measures. We convened key stakeholders in physician performance measurement, documented stakeholder perspectives, and developed a consensus guideline concerning the use and operationalization of exceptions in national physician performance measures.

Completion of these aims involved many of the same parties that worked together for over a year implementing Phase I of Cardio-HIT: 5 practice sites, each with at least 4 years of experience with EHRs; 2 leading performance measure developers—the AMA-convened PCPI and the National Committee for Quality Assurance (NCQA); and the Iowa Foundation for Medical Care (IFMC), the Quality Improvement Organization (QIO) responsible for maintaining national databases for the CMS QIO program and experienced in data abstraction from both paper and electronic medical records.

Scope

Background on Exception Reporting

Definition of Exceptions and Issues for Consideration. In the context of physician performance measures, *exceptions* (also known as *exclusions*) refer to valid reasons why a physician may remove patients from the denominator of measures to calculate performance. The participating physicians recorded data on exceptions prospectively (ie, documentation of exceptions is a routine part of patient management). Broad exception categories (medical, patient, system) are specified for each PCPI measure based on the clinical appropriateness of the exception type for the measure. Some measures have no valid exception categories while others have all three. To illustrate with a measure regarding the prescription of a medication, a clinical contraindication and medical reason for an exception would be an allergy. Whereas, assessing an osteoarthritis patient's satisfaction with function and pain, it is unlikely that there would be any valid medical, patient or system reason as to why a physician could not complete a patient assessment. In this way, physicians have the opportunity to "account for" 100% of their patient

population for each measure. Some would argue that this method helps to avoid patients "falling through the cracks," while others posit that it encourages "loopholes."

Function of Exceptions. Exceptions are an element of physician performance measure design that are intended to fulfill 4 functions: promote appropriateness of care, facilitate quality improvement and patient management, track variations, and prevent unintended penalization of physicians.

Appropriateness of Care. Process-based performance measures can potentially create incentives for physicians to adhere to guidelines for care that are implicit in the measures without appropriate consideration of clinical, patient or situational factors that may warrant deviation from standards.[2] Such an unintended and paradoxical consequence would contradict the underlying philosophy of performance measurement as a tool for promoting and facilitating evidence-based, patient-centered care.[3] Allowing physicians to exclude patients for whom care measured by a given indicator would be clinically inappropriate or in violation of a patient's wishes preserves the priority of physician decision-making and the physician-patient relationship over blind conformity to guidelines and measurement standards.[4]

Quality Measurement as a Tool for Quality Improvement and Patient Management. Beyond serving as a means for provider feedback, performance measures are intended to serve as tools for patient management and quality improvement[5,6] Performance measure definitions and specifications are being increasingly integrated within decision support programs and data collection instruments. Integrating exception criteria within patient management and HIT yields several potential benefits: physicians are reminded of guideline criteria to support appropriate care at the point of care, unnecessarily redundant data collection may be minimized, and thorough history taking may be encouraged. Moreover, periodic review and reconsideration of patients identified as having valid exceptions may be equally important to identifying patients who did not receive the treatment and do not have a valid exception.

Unintended Penalization of Physicians. Performance measurement can potentially penalize physicians who treat disproportionate shares of severely ill patients or difficult-to-treat/reach populations.[7,8] A central tenet of fair, accurate performance measurement is adjustment for heterogeneous patient risks.[9,10,11,12] Full, clinical risk adjustment of performance measurements is infeasible on a broad scale due to its data- and technically-demanding nature.[13,14] Moreover, clinical risk adjustment models tend to be valid within narrowly defined scopes: developing and validating risk adjustment models for a large number of conditions to be applied to a heterogeneous patient and provider population, and which must be subject to regular revision; would be prohibitively costly. Thus, performance measurement systems with anticipated large-scale application (such as the PCPI) have opted to incorporate exceptions within the design of their measures as a parsimonious alternative to risk adjustment.

Context: Site Setting Descriptions and Participants

Overview. The Cardio-HIT Phase II collaborative is a research network of 5 independent physician group practices that was formed to demonstrate the feasibility and value of implementing the national performance measures for CAD and HF within existing EHRs. The 5

groups comprising the Cardio-HIT Collaborative are: Fox Prairie Medical Group (FPMG); Midwest Heart Specialists (MHS); North Ohio Heart Center (NOHC); Physicians Health Alliance (PHA); and University of Pittsburgh Medical Center (UPMC). These groups are located in Illinois, Ohio, and Pennsylvania, and vary in a number of practice characteristics, including: specialty, academic affiliation, physician group size, patient volume, and organizational structure. Although all practices have EHRs in use, they vary in the specific EHRs product being used. Five different EHRs products are currently in use by the Cardio-HIT sites: Epic (UPMC); NextGen® (FPMG); Cardioworks® (MHS), Touchworks by Allscripts® (NOHC), and GE Centricity® (PHA). Experience with their EHRs varies across the 5 sites from 4 years to over 10 years. With the exception of Cardioworks, which is a specialized product for cardiology practices, the other 4 products rank among the market leaders. Although each installation of a product differs, our results provide instrumental insights to other users of these major EHRs products, and thus reflect a relatively large proportion of the segment of ambulatory medical care practices that have adopted EHRs. Additional details on the Cardio-HIT practices are described below.

Table 1. Description of Cardio-HIT sites

Cardio-HIT Site	EHRs Product	Years of EHRs Use	Total # MDs in Practice Group	# Office Locations	Specialties in Cardio-HIT
MHS	CardioWorks	10	53	6 Main, 11 Satellite	Cardiology
UPMC	Epic	7	83 (45 Faculty, 38 Residents)	1	General Internal Medicine
NOHC	Allscripts	10	31	8	Cardiology
PHA	GE Centricity	8	46 (17 in Cardio-HIT)	8	Internal Medicine/ Family Practice
FPMG	NextGen	4	3	1	General Internal Medicine

Methods

Study Design

A 2-year observational study accomplishing 3 aims:

Specific Aim 1. Develop an empirical understanding about prevalence and patterns of exception reporting among physicians using EHRs and reporting national performance measurements. Exception and performance reporting data were used from the Cardio-HIT sites to quantify prevalence and patterns of exceptions and performance for 2 measure sets: CAD and HF.

Specific Aim 2. Evaluate the feasibility and accuracy of exception reporting among physicians. To achieve this aim, organizational evaluations were conducted to (a) characterize and assess the ability of EHRs-enabled practices to capture data required for exception reporting,

and to assess variation in this process; and (b) evaluate the accuracy and validity of automated exception reports and identify key sources of measurement error.

Specific Aim 3. Analyze and then address stakeholder perspectives concerning exception reporting in physician performance measurement to develop refined principles and methods regarding the use of exception reporting in performance measures.

We convened key stakeholders in physician performance measurement, documented stakeholder perspectives, and developed a consensus guideline concerning the use and operationalization of exceptions in national physician performance measures.

The study built on existing research and technological infrastructures developed in the original Cardio-HIT Phase I project to measure physician performance in CAD and HF using the ACC/AHA/PCPI measure sets. The proposed study expanded existing Cardio-HIT data collection and measurement activities to include collection, measurement, analysis, validation of exception data and focus groups with key stakeholders.

Data Sources/Collection

Overview. Two sources of data were used for this study: PCPI physician performance measure data, which were collected by all Cardio-HIT sites; and detailed data on reported performance and exceptions, which were collected through explicit EHRs medical record abstraction. The process of integrating the ACC/AHA/PCPI measures into an EHRs, importing de-identified data to a central warehouse, and the subsequent report development and distribution was a multi-phased approach involving each practice and IFMC, which managed the Cardio-HIT clinical data warehouse. The team from the individual practices consisted of staff that was well versed in the technical requirements of data extraction from their EHRs as well as staff that possessed detailed knowledge of the clinical workflow of the practice as it relates to how clinicians are utilizing the EHRs data fields.

Collection of Physician Performance Data (PCPI Performance Measures). The integration of the performance measures involved the translation and mapping of the measure specifications (numerators, denominators and exceptions) into the specific data fields available within each practice EHRs. As each practice had a unique set of data fields, this step required individual mapping of the data elements at the practice level. Next, performance measurement data were extracted from the practice EHRs for the time period under study. An interface template was developed for each practice EHRs, which contained the unique set of data fields, validation requirements and acceptable values associated with ACC/AHA/PCPI measures. Using the interface template, each practice queried its EHRs database to compile the data elements required for each measure. To assure consistent capture of data across a diverse set of EHRs systems, the interface template identified the submission of the prescribed coding system or standardized medical vocabulary as defined per the ACC/AHA/PCPI measure.

In cases where the EHRs were unable to support a standard coded medical vocabulary, an acceptable alternative was to substitute a Y/N value for those fields. For example, none of the EHRs included RxNorm coding for medications; some sites used NDC codes but other sites used their own formulary ID number or text to determine if a patient was on a specific medication. In

the latter case, the site sent a Y/N value for the measure, with the intent to migrate to Health Information Technology Standards (HITSP) in the future.

Once data were extracted from the EHRs system, the practice accessed the secure portal and uploaded its de-identified, patient dataset. This dataset underwent an extensive import processing routine that validated each field in accordance with the permitted values, or range of values, as defined within each clinical measure. Based on these validation routines, any data errors were reported on the Submission Reports that enabled the practice to correct the situation and resubmit its data.

Following successful submission of data, measurement reports were generated that processed the results of each measure submitted by the practice. These reports list the denominator, numerator, and exceptions applicable to each measure. Reports were summarized by physician, by clinic, and by the overall practice.

Collection of Exception Validation Data. The purpose of validation was to verify the accuracy and/or completeness of data used for calculation of physician performance measurement. In this case, we validated the exceptions collected and submitted to the Cardio-HIT warehouse by the 5 sites. We designed an onsite validation process for the office practice sites that included the following: (1) specification of a timeframe for validation; (2) specification of measures and tools used for validation; (3) identification of the universe of the physician practices participating; and (4) creation of a sampling methodology using the identified physician practices. The data-gathering process included: (1) accessing the EHRs for the sampled practices; and (2) analyzing clinical data to obtain a comparison of the clinical record to the data that was submitted by the practices to the Cardio-HIT warehouse. Skilled and experienced IFMC audit staff traveled to each site to conduct the audit. We created a validation monitoring report that calculated the percentage of agreement between the submitted (to the warehouse) and abstracted data.

Sampling. For this analysis, we drew random samples from all reported exceptions and all apparent quality failures, yielding 2 samples for separate analyses: an exception sample for validation of exception criteria; and an apparent quality failure sample for validation of apparent quality failures. A third abstraction sample for the HF data was drawn and analyzed to determine the sensitivity and specificity of the practice site performance measure results (measure met or positive performance sample). Sample sizes were determined from statistical power calculations, using 80% power, a 5% margin of error, and a pre-study conservative estimate of 50% population proportion. The calculation was performed separately on the number of measure failures and the number of exceptions from the measure, as different questions were studied from these 2 groups. For the exceptions, the sample size allowed a probability of 80% that the estimate of the proportion of exceptions that were valid exceptions would be no farther than 5% from the proportion of exceptions that were valid in the entire population. For the measure failures, the sample size allowed a probability of 80% that the estimate of the proportion of identified failures that were true failures was within 5% of the same proportion among the entire population. Cases (patients) were sampled randomly.

Abstraction Tool. Electronic medical records for patients in the exception, apparent quality failure and measure met samples were accessed for abstraction using an explicit tool designed by the team. This abstraction tool was specifically designed to extract information from a medical

chart to: (1) determine the existence of physician chart documentation of reason(s) for exception; (2) the specific reason for exception; and 3) to determine the sensitivity and specificity of the practice site performance measure results. The Cardio-HIT Clinical Director with the Physician Co-Investigators constructed an a priori list of "valid" or "acceptable" medical, patient, and system reasons for each of the 7 measures with explicit allowances for exception. Two, trained abstractors conducted on-site searches, reviews for exceptions, and determined whether each reported exception was valid based on the a priori list of acceptable reasons for exception. When disagreement occurred between the auditors or when a documented reason in the record was not included in the list, the site Physician Co-investigator was consulted.

Measures

The performance measures used in this project focused on CAD and HF. They were developed by the ACC, the AHA, and the PCPI. To date, the PCPI has developed over 250 performance measures and is comprised of representatives from more than 170 member organizations including national medical specialty and state medical societies, the Council of Medical Specialty Societies, American Board of Medical Specialties and Member Boards, experts in methodology and data collection, the Agency for Healthcare Research and Quality and CMS.

National recognition is increasing for these performance measures. As of March 2010, the National Quality Forum has endorsed more than 100 PCPI performance measures for ambulatory care, including the 7 ACC/AHA/PCPI CAD and HF performance measures listed below. The AQA has selected these same measures as appropriate for measuring physician performance. The CAD and HF performance measures have been implemented into three CMS demonstration projects: Medicare Care Management Performance, EHR, and the Physician Group Practice. Several of these measures are included in PQRI for physician reporting and the proposed meaningful use and adoption of EHRs rule. The following measures have one or more exceptions and were our focus:

- CAD: Antiplatelet Therapy
- CAD: Drug Therapy for Lowering LDL-Cholesterol
- CAD: Beta-blocker Therapy – Prior Myocardial Infarction
- CAD: ACE Inhibitor/ARB Therapy
- HF: Beta-blocker Therapy
- HF: ACE Inhibitor/ARB Therapy
- HF: Warfarin Therapy for Patients with Atrial Fibrillation

Detailed specifications for these measures for use in EHRs have been developed and are available at www.physicianconsortium.org.

Study Limitations and Strengths/Innovations

Study Limitations. An aspect of our study that limits the generalization of our results is the non-random selection of EHRs-enabled study sites. In our case, site self-selection into the original Cardio-HIT project may have resulted in a set of sites that are more adept in EHRs use and which are more interested and motivated with respect to performance measurement and quality improvement. The empirical picture that emerges from our study with respect to the use of exceptions may reflect a best case scenario. The prevalence and patterns of exception reporting among a random sample of EHRs-enabled practices or among a sample of non-EHRs practices who use PCPI performance measures may be considerably different.

A second limitation of our study was the restriction of validation to apparent quality failures as this reflects an implicit, untested assumption that the data and identification of "quality successes" (ie, the cases identified for the numerator of performance measures) are accurate. A comprehensive study would subject this assumption to equally rigorous testing. Due to resource constraints as well as the pressing need to address the problem of exception reporting in the context of national health policy debates, we have chosen to place priority on evaluating the accuracy and validity of apparent quality failures.

A third limitation of our study was that the measures under study were limited to selected PCPI CAD and HF measures with exception provisions all pertaining to medications. Process measures of physician performance focusing on other behavioral forms of physician care in which patient or system reasons for exception are predominant may yield different patterns and prevalence of exceptions. One may expect that, in a prospective measurement system, process measures where patient and system reasons are particularly important may be those measures where opportunistic reporting may be most evident. As an illustration, short of interviewing patients, there is no way to independently validate physician documentation of patient reasons for exception.

Study Strengths and Innovations. In addition to the significance of the proposed research, we note additional strengths and innovations of our study. First, ours was a multi-site, multi-EHRs product study. This design enabled us to begin assessing the extent to which variation in prevalence and patterns of exception reporting are due to EHRs idiosyncrasies and measurement issues, rather than due to physician behavior. Second, we innovated and developed an explicit abstraction tool for collecting data to validate exceptions for 2 national physician performance measure sets—the ACC/AHA/PCPI CAD and HF measure sets. Third, we demonstrated the explicit collection and validation of exception data on a larger and more comprehensive scale than previous efforts,[15,16] and may set empirical precedence for future research and regulatory work by PQRI, for example. Fourth, we explored the potential application of instrumental variables methods for testing one mechanism of physician gaming in performance measurement and reporting to address the endogeneity bias that inheres in the empirical question.[17]

Results

Principal Findings

We analyzed patient-level data from the five practices participating in the Cardio-HIT quality measurement/EHR collaboration project for the above noted AHA/ACC/PCPI CAD and HF performance measures.

CAD Findings. The practices reported data on 47,075 CAD patients for four CAD drug therapy performance measures (Antiplatelet, LDL lowering, Beta-blocker, and ACEI/ARB), including exception reasons. Retrospective manual reviews of the EHRs were conducted on a sample of 538 patients with reported exceptions.

- Among patients with reported exceptions, there was 93% (95% CI: 90.3%, 94.9%) agreement between the reported exception and documentation in the EHRs based on an a priori list of appropriate exceptions.

- The "true exception" rate where an exception was reported and no drug was prescribed, across all sites and all measures was 3.5%, with variation across the measures ranging from 2.0% to 6.2%.

- Majority of records with reported exceptions, 74.6%, also met the numerator due to the medication being prescribed.

- The location of exceptions in the EHR and whether it was in coded form varied by measure and EHRs.

Additional CAD findings include:

- Overall performance rate for all sites for the four CAD measures was 76.7%.

- Of the 167 patients with true exceptions, 97.5% (95% CI: 94.6%, 100.0%) had a reported exception found to be in agreement with the exception in the EHRs documentation and the a priori list.

- Majority of true exceptions, 98.6%, were medical reasons.

- In the case of true exceptions, across the four measures, there were specific medical reasons found in the EHRs:

    - clinical contraindication - 63.0%;

    - drug intolerance - 19.3%;

    - drug allergy - 16.2%; and

- drug interaction - 1.4%.

- Among patients with "apparent quality failures," 25.4% (95% CI: 21.8%, 29.0%) were found to actually be opportunities for improvement.

HF Findings. Practices reported to the data warehouse 13,985 eligible HF patients for three HF drug therapy performance measures (Beta-blocker, ACEI/ARB, and Warfarin), including exception reasons. Retrospective manual reviews of the EHRs were conducted on a sample of 559 patients with exceptions reported to the data warehouse that included patients with multiple exceptions and patients who met the numerator.

- Among patients with reported exceptions, there was 87% (95% CI: 83.2%, 91.5%) agreement between the reported exception and documentation in the EHRs based on an a priori list of appropriate exceptions.

- The "true exception" rate where an exception was reported and no drug was prescribed, across all sites and all measures was 5.6%, with variation across the measures ranging from 5.3% to 6.2%.

- Majority of records with reported exceptions, 77.7%, also met the numerator due to the medication being prescribed.

- The location of exceptions in the EHR and whether it was in coded form varied by measure and EHRs.

Additional HF findings include:

- Overall performance rate for all sites for the three HF measures was 76.8%.

- Of the 86 patients with true exceptions, 95.3% (95% CI: 90.3%, 100.0%) had a reported exception found to be in agreement with the exception in the EHRs documentation and the a priori list

- Majority of true exceptions, 97.2%, were medical reasons.

- In the case of true exceptions, across the three measures, there were specific medical reasons found in the EHRs:

    - clinical contraindication - 85.6%;

    - drug intolerance – 5.7%;

    - drug allergy – 8.7%; and

    - drug interaction - 0.0%.

- Among patients with "apparent quality failures," 22.4% (95% CI: 17.6%, 27.9%) were found to actually be opportunities for improvement.

Measure Met Findings. Practices reported to the data warehouse 12,403 eligible HF records for six HF performance measures (Beta-blocker, ACEI/ARB, Warfarin, Left Ventricular Ejection Fraction Assessment, Weight Measurement and Blood Pressure Measurement). Retrospective manual reviews of the EHRs were conducted on a sample of 678 records reviewed for cases where the numerator was documented and an exception was identified in the EHR review.

- Majority of records, 85.1%, were reported as having met the measure as the numerator was documented and no exception was identified.

- Few records, 0.7%, were identified as "misclassification – exception found" because the numerator was reported to the warehouse, which upon manual abstraction of the EHR, the measure was found not to be met and an acceptable exception was identified.

- A small number of records, 14.2%, were identified as "invalid – apparent quality failures" because the numerator was reported to the warehouse, which upon manual abstraction of the EHR, the measure was found not to have been met and an acceptable exception was not identified.

Discussion

Our data indicate that similar to the United Kingdom study conducted in 2008 by Doran and colleagues regarding the exclusion of patients from pay-for-performance targets by English physicians the overall, applied exception rates were modest. The rate for each quality measure was lower than the rate reported for similar measures used in the United Kingdom, particularly for beta-blocker therapy for prior MI, where the Cardio-HIT exception rate was 6.1%, and the U.K. rate was 25.3%.[18] Some of this variation may be attributed to differences in measure specifications. In addition, the Cardio-HIT practice sites, unlike practices in the U.K., had limited experience with quality measure reporting and had no financial incentives for quality reporting. For the two measures also included in the CMS 2007 PQRI, the Cardio-HIT exception rates were lower as well (2.0% vs. 4.2% for antiplatelet therapy and 6.2% vs. 8.1% for beta-blocker therapy for prior MI).

The high rates of agreement found when validating the reported exceptions against manual review support the use of exceptions in quality reporting. Our study did not show evidence of misuse of exceptions. For these measures, quality failures would have been overestimated were reporting of exceptions not allowed. However, physicians, measure developers, and EHR vendors must work together to improve automatic querying and reporting of exceptions.

As our results revealed, exception rates will vary depending on whether exceptions *not* applied are included. The methodology utilized by the PCPI and the U.K. results in reporting applied exceptions only, which seems appropriate for performance reporting. Moreover, documenting applied exceptions in an EHR is important for care coordination, signaling to another provider why a drug was not prescribed. We believe that the cases where an exception was noted, but the physician prescribed the drug, demonstrate that most exceptions are not absolute. Performance measures are not intended to replace careful decision-making.[19,20] For

example, although myalgia is an acceptable medical exception for drug treatment for lowering LDL-C, the physician and patient may decide to continue with a drug, depending on the myalgia severity and CAD risk, or opt for another drug from the same class.

We present more granular information on medical exceptions than that previously reported from the United Kingdom. We suggest subcategorizing medical exceptions for these types of measures into clinical contraindication, drug allergy, drug interaction, and drug intolerance. External reporting across these subcategories—if accomplished without high data entry burden and on a broad scale—may provide valuable information to guideline developers, pharmaceutical manufacturers, and measure developers. We were not able to further evaluate patient and system reasons for exceptions because they were reported infrequently for these measures.

Our finding that approximately 75% of cases considered apparent quality failures by automatic query and reporting were in fact cases where the measure was met is consistent with previous work that found 15% to 81% misclassification of apparent quality failures in automatic EHR reporting.[21] Potential reasons for misclassifications include problems with drug codes (eg, multiple coding schemes in the EHRs, frequency of drug code updates, difficulty maintaining up-to-date drug lists in the measure specifications, lack of documentation of over-the-counter aspirin use). The recent proposal to utilize RxNorm as the vocabulary for drugs in quality measures may address these issues.[22] Also, during manual review, abstractors searched for exceptions for up to three years prior to data reporting and searched in multiple areas of the chart including free text, problem lists, and allergy lists, whereas the automated reporting was based on a one-year look back (ie, third quarter 2006 through second quarter 2007). The currency of data within EHRs is likely to increase with focused attention on quality reporting from EHRs.

Lastly, we found variation in the location of data pertinent to exceptions in the patient EHRs and found that most of the data were not in coded form. Ongoing discussions among measure developers, EHR vendors, and EHR users,[23] as well as the evolution of certification of EHR products, may facilitate the availability of coded data within EHRs for quality measurement.

Conclusions

Exception reporting was generally low with high rates of agreement and identified aspects of care important to capture in an EHR for care coordination and patient safety. The specific reasons for a medical exception suggest standard categories of medical exceptions (eg, clinical contraindications, drug allergy or interaction). Because many exceptions are not absolute, physicians may decide to "over-ride" an exception and provide the relevant aspect of care. Automatic reporting often missed critical information. With improved automation, additional granularity may be possible.

Physicians are more likely to accept the quality measure results as valid if they can account for exceptions. Others report physician frustration when exceptions are not permitted.[24] Exceptions also provide a means to track variations in care and focus quality improvement efforts. For example, rather than simply reporting that 40% of eligible patients did not receive a particular drug, the data can show that 30% did not receive the drug for a reported medical reason and 10% did not receive the drug with no reason provided. Further investigation of the 10% may be a good first step in targeting quality improvement efforts. On the other hand, concerns exist that exception reporting will be used excessively. A suggested middle ground is to allow exception reporting but monitor its use.[25]

Significance and Implications

The ability to collect and analyze exception data may prove valuable in understanding variations in care. Physician access to exception data from the EHRs at the point of care is critical for decision-making and may help improve patient outcomes, perhaps through clinical decision support systems. If widely available, payers and policy makers could reliably use physician performance measures reported from EHRs for quality reporting, pay for performance, and identifying outlier rates in larger, population-based measurement programs to target reasons and variations among patient populations. These findings will enable the development and dissemination of health IT evidence and evidence-based tools to improve health care decision-making using integrated data and knowledge management.

References

1. Epstein AM. Performance reports on quality – prototypes, problems and prospects. *N Engl J Med*. 1995;333:57-61.

2. Walter LC, Davidowitz NP, Heineken PA, Covinsky KE. Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA Performance Measure. *JAMA*. 2004;291:2466-70.

3. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone, 2000.

4. Casalino LO. The unintended consequences of measuring quality on the quality of medical care. *N Engl J Med*. 1999;341:1147-1150.

5. Casalino LO. The unintended consequences of measuring quality on the quality of medical care. *N Engl J Med*. 1999;341:1147-1150.

6. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. *JAMA*. 2000;283:1866-1874.

7. Epstein AM. Performance reports on quality – prototypes, problems and prospects. *N Engl J Med*. 1995;333:57-61.

8. Zaslavsky AM. Statistical issues in reporting quality data: small samples and case mix variation. *Int J Qual Health Care*. 2001;13(6):481-488.

9. Weiss KB, Wagner R. Performance measurement through audit feedback and profiling as tools for improving clinical care. *Chest*. 2000;118:53-58.

10. Epstein AM. Performance reports on quality – prototypes, problems and prospects. *N Engl J Med*. 1995;333:57-61.

11. Dranove D, Kessler D, McClellan M, Satterthwaite M. Is more information better? The effects of "report cards" on health care providers. *J Polit Econ*. 2003;111:555-588.

12. Kassirer JP. The use and misuse of practice profiles. *N Engl J Med*. 1994;330:634-636.

13. Mukamel DB, Mushlin AI. The impact of quality report cards on choice of physicians, hospitals, and HMOs: a midcourse evaluation. *Joint Comm J Qual IM*. 2001;27:20-27.

14. Epstein AM. Performance reports on quality – prototypes, problems and prospects. *N Engl J Med*. 1995;333:57-61.

15. 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;166:2272-2277.

16. 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.

17. Doran T, Fullwood C, Reeves D, et al. Exclusion of patients from pay-for-performance targets by English physicians. *N Engl J Med*. 2008;359:274-84.

18. Doran T, Fullwood C, Reeves D, et al. Exclusion of patients from pay-for-performance targets by English physicians. *N Engl J Med*. 2008;359:274-84.

19. Antman EM, Peterson ED. Tools for guiding clinical practice from the American Heart Association and the American College of Cardiology: what are they and how should clinicians use them? *Circulation*. 2009;119:1180-1185.

20. Garber AM. Evidence-based guidelines as a foundation for performance incentives. *Health Aff*. 2005; 24:174-179.

21. 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.

22. Office of the National Coordinator HIT Standards Committee Meeting. Presentation, Washington, D.C. August 20, 2009.

23. American Medical Association/HIMSS Electronic Health Record Association /National Committee for Quality Assurance. Collaborative for Performance Measure Integration with EHR Systems. Available at: http://www.ama-assn.org/ama/pub/ physician-resources/clinical-practice-improvement/programs-events/collaborative-performance-measure-integration-ehr-systems.shtml. Accessed July 1, 2009.

24. McDonald R, Roland M. Pay for performance in primary care in England and California: comparison of unintended consequences. *Ann Fam Med*. 2009;7:121-127.

25. Conway PH, Clancy C. Transformation of health care at the front line. *JAMA*. 2009;301(7):763-765.

List of Publications and Products

Pending Publications

Reporting on National Clinical Performance Measures from Ambulatory Electronic Health Records: Lessons Learned from the Cardio-HIT Project (Phase I)

Reporting Cardiovascular Performance Measures from Electronic Health Records: An Analysis of Exception Reporting from the Cardio-HIT Project (Phase II)

In Development

Reporting Heart Failure Performance Measures from Electronic Health Records: An Analysis of Exception Reporting from the Cardio-HIT Project (Phase II)