

Electronic Health Records and the Reliability and Validity of Quality Measures: A Review of the Literature

Medical Care Research and Review
67(5) 503–527
© The Author(s) 2010
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/1077558709359007
<http://mcr.sagepub.com>



Kitty S. Chan,¹ Jinnet B. Fowles,² and Jonathan P. Weiner¹

Abstract

Previous reviews of research on electronic health record (EHR) data quality have not focused on the needs of quality measurement. The authors reviewed empirical studies of EHR data quality, published from January 2004, with an emphasis on data attributes relevant to quality measurement. Many of the 35 studies reviewed examined multiple aspects of data quality. Sixty-six percent evaluated data accuracy, 57% data completeness, and 23% data comparability. The diversity in data element, study setting, population, health condition, and EHR system studied within this body of literature made drawing specific conclusions regarding EHR data quality challenging. Future research should focus on the quality of data from specific EHR components and important data attributes for quality measurement such as granularity, timeliness, and comparability. Finally, factors associated with poor or variability in data quality need to be better understood and effective interventions developed.

Keywords

electronic health record, quality of care, measurement, reliability, validity

This article, submitted to *Medical Care Research and Review* on March 30, 2009, was revised and accepted for publication on December 3, 2009.

¹Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

²Park Nicollet Institute, Minneapolis, MN

Corresponding Author:

Kitty S. Chan, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway, Room No. 633, Baltimore, MD 21205-1901

Email: kchan@jhsph.edu

Introduction

Quality measures have become an integral part of the U.S. health care system. They are used to monitor quality of care, assess the effectiveness of quality improvement initiatives, and serve as the basis for pay-for-performance programs and public accountability reports (Institute of Medicine, 2006). These measures have traditionally been derived from manual reviews of medical records, secondary analysis of billing or administrative databases, or via direct surveys of patients or consumers. However, as electronic health records (EHRs) diffuse throughout the U.S. health care system (desRoches et al., 2008), new electronic functions, such as problem and medication lists, decision-support modules, and patient input via Web portals, will create opportunities and challenges in terms of data availability, quality, and comparability (Fishman, Hornbrook, Meenan, & Goodman, 2004). In this article, we reviewed evidence from empirical studies on EHR data quality published in the past 5 years, with an emphasis on data attributes that are important to quality measurement. Based on these findings, we identify gaps in the literature and suggest directions for future research. Furthermore, we propose roles that health care organizations and third party regulatory or accreditation organizations should play to enhance the value of EHR data for quality measurement.

New Contribution

Thiru, Hassey, and Sullivan (2003) conducted a systematic review of studies on the quality of EHR data for primary care. While informative, their review did not address the specific data needs of different applications of EHR data, such as quality measurement. Furthermore, since Thiru et al. (2003) examined studies published from 1980 to 2001, recent studies reflecting the rapid changes to technology and the current health care environment were not represented. To address these gaps, this review will focus on peer-reviewed studies published since 2004 and data attributes that have important implications for quality measurement.

Conceptual Framework

The goal of this article is to examine how the reliability and validity of quality of care measures may be influenced by the evolving electronic data environment, particularly the emergence of the EHR. Data completeness and accuracy have been used in previous studies to describe data quality (Pringle, Ward, & Chilvers, 1995; Thiru et al., 2003; Williams, 2003). Data completeness is defined as the level of missing data for a data element. Data accuracy is conceptualized as the extent to which data captured through the EHR system accurately reflects an underlying state of interest (e.g., whether a medication list accurately reflects the number, dose, and specific drugs a patient is currently taking). This includes the timeliness or "currency" (Williams, 2003) of the data as well as the clinical specificity or "granularity" in how individual clinicians or

clinical staff document within an EHR system. A high level of missing data and/or data errors will substantially reduce measure reliability and validity. Although we could not find a generally accepted definition of data comparability, its importance to valid care quality comparisons and outcomes research has been recognized (Dean et al., 2009; Schneider, Riehl, Courte-Wienecke, Eddy, & Sennett, 1999). In this article, we define data comparability as similarity in data quality and availability for specific data elements used in a measure across different entities, such as health plans or physicians or data sources. Variations in EHR content, structure, and data format as well as local data capture and extraction procedures can significantly affect data comparability, as can EHR systems using coding systems or menu choices of varying granularity (e.g., for diagnoses or medications), different data extraction strategies (e.g., frequency of data extraction, specific algorithm used to extract data from structured or free-text fields), or having different data quality standards.

Wherever possible, we organize study findings for each data attribute by specific data element, major EHR functions (e.g., problem or medication list, computerized physician order entry; desRoches et al., 2008), and the data flow process (data documentation, data abstraction, data aggregation across systems). Finally, we consider how the EHR may influence measure validity, including EHR-based specification modifications to enhance reliability and validity.

Method

We searched PubMed for peer-reviewed articles published between January 2004 and June 2009. We used the following search terms: *computerized medical record system, electronic medical record, electronic health record, electronic patient record, quality indicator, performance measure, quality measure, quality measurement, quality of care assessment, quality assurance, health care, quality of health care, data quality, data accuracy, data reliability, data validity, missing data, and administrative data*. Using these terms, we conducted several searches. The first focused on articles related to EHRs, quality measures, and data quality. In this search, we included both U.S. and non-U.S. studies because of the earlier EHR implementation and better developed data quality research in countries such as the United Kingdom (Thiru et al., 2003). The second search focused on articles related to EHRs and quality measures more generally. Finally, we searched for articles related to EHRs and administrative data, given the role of administrative data in quality measurement. For the latter two searches, we focused on U.S. studies because we are particularly interested in quality measurement within the U.S. context. For all searches, we excluded studies conducted solely in the inpatient or institutional (e.g., nursing home) settings. We also excluded duplicates, news, tutorials, editorials, or other articles that did not include empirical data.

From these searches, we identified and retrieved the abstracts for 532 articles. Using our inclusion and exclusion criteria, we retained and retrieved the full text for 101 articles. We used the same criteria to exclude another 66 articles after full text review. Of the remaining 35 articles, we abstracted information on study period, study

design, data or measure attribute(s) assessed, context (U.S., non-U.S., integrated delivery system, single clinic, clinical network, etc.), clinical condition, EHR component or data flow step examined, data elements, and study results.

Results

Of the 35 articles (see Table 1), 28 were conducted in the United States and 7 were conducted outside the United States. Studies were diverse with respect to study settings, health issues, and populations. Some studies focused on a specific EHR component, such as the medication list, but were nonspecific with respect to health condition or population (Kaboli, McClimon, Hoth, & Barnett, 2004; Staroselsky et al., 2008; Varkey, Cunningham, & Bisping, 2007). Others examined the general accuracy of EHR data without identifying the particular data elements or EHR components under evaluation (Powell, Fitton, & Fitton, 2006). Studies used relatively weak comparative ($N = 22$) or observational descriptive ($N = 8$) study designs. In the absence of a “gold standard,” less reliable reference standards such as manual chart abstractions, surveys, or interviews are used to evaluate EHR data quality. Five studies were controlled trials or pre-post evaluation of data quality interventions. Many studies evaluated multiple aspects of data quality or comparability, with 23 studies (66%) evaluating data accuracy, 20 studies (57%) data completeness, 8 studies (23%) data comparability across study sites or data systems, and 3 studies (9%) on measure validity. Few studies explicitly addressed data granularity or data timeliness.

Data Completeness

Studies of data completeness are characterized by diversity in the study settings, populations, and data elements examined. Preliminary evidence suggests that data quality within medication and problem lists were mixed or poor. Other EHR components were rarely studied. Local data capture, which includes autopopulation of data through electronic linkage with outside sources and manual entry into structured or free-text data fields, and data extraction practices exert significant influence on data completeness. The findings from reviewed studies are presented below by data element content, EHR component, and data flow process.

Data element content. Data completeness for different data elements, including clinical status, laboratory results, and clinical actions or procedures, was most often reported within single studies and generally varied widely (Table 2). The EHR data sources for these elements were often not clearly specified, although manually entered structured data fields are likely to be a key source. Among data elements that were examined by more than one study, data completeness varied substantially across studies, ranging between 0.1% and 51% for blood pressure and 10% and 38% for smoking status assessment. Clinical relevance could be an important contributor to data completeness. Persell, Dunne, Lloyd-Jones, and Baker (2009) found less missing data for blood pressure and cholesterol values in a high cardiac risk group compared with the general

Table 1. Reviewed Articles (N = 35)

Article Reference (Study Period)	Attribute Examined	Context	Study Design	Health Condition/ Patient Population ^a	EHR Component or Data Flow	Data Elements
Agnew-Blais et al., 2009 (2005)	Completeness	U.S., Brigham and Women's Hospital Arthritis Center	Comparative: EHR data vs. doctor's medical record notes ("reference")	Rheumatoid arthritis; 100 randomly selected outpatients	Data documentation; medication list	Prescription for DMARD
Linder et al., 2009 (2004-2005)	Completeness; comparability	U.S., adult patient from 8 primary care clinics in Greater Boston	Descriptive	Community- acquired bacterial pneumonia	Problem list; structured fields	Health conditions, smoking status, smoking cessation counseling, flu immunization status, vital signs, oxygen saturation, mental status, hydration, chest radiograph, level of care rationale, antibiotic appropriateness
McGinnis et al., 2009 (2002)	Completeness; accuracy comparability	U.S., VA (national data)	Comparative: 2 data repositories (DSS and ICR)	HIV; Veterans with visit (N = 22,647)	Data extraction	Hemoglobin, aspartate aminotransferase, alanine aminotransferase, glycosylated hemoglobin, creatinine, white blood cell, glucose, HIV RNA viral load, CD4+ count
Persell, Dunne, et al., 2009 (2005-2006)	Measure validity	U.S., single large urban academic internal medicine practice	Comparative: conventional vs. more complex EHR-based measures	Hypertension; 5,905 adults (18+ years) with 3 or more clinic visits	Structured fields	Diagnosis, BP
Persell, Kho, et al., 2009 (2004-2006)	Completeness; accuracy	U.S., single large urban academic internal medicine practice	Comparative: EHR structured data vs. manual review ("reference")	Cardiac risk; (N = 23,111 adult patients; N = 400, 100 in each risk group)	Automated risk classification	Age, gender, systolic BP, mean total and HDL cholesterol, use of antihypertensive, lipid-lowering, antithrombotic medications, smoking status
Baldwin, 2008 (2001-2003)	Accuracy	U.S., large academic medical center's women health center	Comparative: natural language processing program vs. manual review ("reference")	Breast cancer prevention; 60 women ≥40 years stratified convenience sample	Data extraction (NLP)	Age, race, sex, risk factors, screening, and test results (exams, treatment, follow-up)

(continued)

Table 1. (continued)

Article Reference (Study Period)	Attribute Examined	Context	Study Design	Health Condition/ Patient Population ^a	EHR Component or Data Flow	Data Elements
Forster et al., 2008 (2006-2007)	Completeness	Non-U.S. (resource- limited setting; 15 countries); 21 treatment sites	Descriptive	HIV patients receiving ART; (N = 50,060)	Not specified	Age, gender, year ART initiated, clinical stage, baseline CD4, follow-up CD4
Goodyear-Smith et al., 2008 (2005-2006)	Completeness; accuracy; comparability	Non-U.S. (New Zealand): 124 randomly selected general practices	Descriptive, logic checks	Childhood immunization	Data documentation; structured fields	Vaccine receipt, dates
Pakhomov, Jacobsen, et al., 2008 (2006)	Accuracy	U.S., Mayo Clinic: 12% hospital records, remainder outpatient	Comparative: EHR structured data vs. patient report	Heart disease symptoms; 1,110 randomly selected patient records with one of the symptoms	Data extraction (NLP)	Chest pain, dyspnea, cough
Pakhomov, Bjornsen, et al., 2008 (2000-2004)	Accuracy	U.S., Mayo Clinic primary care clinic	Comparative: EHR structured data vs. manual review ("reference")	Diabetes; 401 randomly selected patients	Data extraction (NLP)	Neurological, vascular, structural components of a foot exam (FE)
Rosenman et al., 2008 (1998- 2002)	Completeness	U.S., Indiana: large public hospital and clinics	Comparative: EHR structured data vs. Medicaid claims data	Prenatal syphilis screening; Medicaid pregnant women (N = 3,960)	Not specified	Diagnosis and procedure codes (claims), dates for all laboratory tests for syphilis from EHR
Staroselsky et al., 2008 (not specified)	Completeness	U.S., single primary care practice	Controlled trial of patient Web portal	Not specified; control N = 79; intervention N = 84	Medication list	Medications, regimen

(continued)

Table 1. (continued)

Article Reference (Study Period)	Attribute Examined	Context	Study Design	Health Condition/ Patient Population ^a	EHR Component or Data Flow	Data Elements
Szpunar et al., 2008 (2002-2003)	Completeness; comparability	U.S., 3 clinic sites from the Henry Ford Health System	Pre-post evaluation of standardized vital sign data entry method	Not specified; 474 randomly selected visits from each clinic	Data documentation; structured fields	Height, weight, temperature, BP, pulse
Baker et al., 2007 (2003-2004)	Completeness; accuracy measure validity	U.S., large urban academic internal medicine practice	Comparative: automated EHR data review vs. automated plus manual review method ("reference")	CHF; N = 517	Data extraction; medical history, problem list, medication list	Diagnoses, prescriptions (beta- blocker, ACE inhibitor, ARB, warfarin)
Goulet et al., 2007 (2001-2004)	Completeness; accuracy; comparability	U.S., VA patients (New England)	Comparative: EHR structured data vs. full chart reviews, External Peer Review Program ("reference")	Hypertension, ischemic heart disease, diabetes	Structured fields	LDL, systolic and diastolic BP, HgA1c
Tang et al., 2007 (2004-2005)	Completeness; accuracy	U.S., single large multispecialty group practice	Comparative: EHR structured data vs. manual review ("reference")	Diabetes; 125 Medicare FFS patients	Data extraction; problem list, structured fields	DMI-DM8 (Doctor's Office Quality project), Diagnoses, procedures, laboratory, and vital sign values
Thielke et al., 2007 (1990-2002)	Accuracy	U.S., VA inpatient and outpatient setting	Descriptive: copying and pasting (at least 40 words of exactly copied text)	Not specified; N = 1,479 randomly selected patients	Data documentation	Clinical exams; Data elements not specified
Varkey et al., 2007 (2005)	Completeness; accuracy	U.S., Mayo Clinic	Preintervention (54 visits, usual care) and postintervention (50 visits, medication reconciliation)	Medication reconciliation; 104 adult primary care visits	Medication list	Medication name, dose, route, and frequency

(continued)

Table 1. (continued)

Article Reference (Study Period)	Attribute Examined	Context	Study Design	Health Condition/ Patient Population ^a	EHR Component or Data Flow	Data Elements
Weingart et al., 2007 (2005)	Completeness; accuracy	U.S., large urban cancer center (Dana Farber)	Pilot pre-post	Ambulatory oncology; N = 338	Medication list	Medication, dose, frequency
Persell et al., 2006 (2004)	Accuracy; measure validity	U.S., large urban academic internal medicine practice	Comparative: automated measurement vs. automated measurement plus manual review for quality failures	CAD; N = 1,006	Data extraction	7 CAD (DOQ-IT) measures; antiplatelet, lipid-lowering, beta-blocker drug prescribed; BP, lipid, LDL control, ACE inhibitor or ARB for diabetics, or left ventricular systolic dysfunction
Powell et al., 2006 (not specified)	Accuracy	Non-U.S. (U.K.) single general practice	Comparative: EHR structured data vs. patient report ("reference")	Not specified; N = 31	Not specified	Not specified
Ricciardi et al., 2005 (2005)	Comparability	U.S., National data warehouse (79 users of commercial EHR)	Descriptive: Data cleaning and terminology mapping harmonization processes	Not specified	Not specified; problem list	Health conditions
Roukema et al., 2006 (not specified)	Completeness	Non-U.S. (Netherlands)	Comparative: Data entry vs. "control" record ("reference")	Pediatric patient history and physical exam	Data documentation	Not specified
Saigh et al., 2006 (not specified)	Accuracy	U.S., Primary care clinics of large urban hospital (Bellevue in NY)	Comparative: Pain Assessment Screen (PAS) vs. manual review ("reference"); pre-post PAS trial	Pain assessment; pre (N = 392) and post (N = 395)	Structured fields	Pain presence, if pain, mandatory: location, severity, optional: frequency, quality, precipitating and alleviating factors and effects on ADL
Staas et al., 2006 (2004)	Accuracy	U.S., single IDS (Intermountain)	Case study: descriptive	Liver transplant recipients; N = 128 (14,082 results from 85 external labs)	Data documentation; structured fields	Date and value of creatinine, cyclosporin A, tacrolimus, absolute neutrophil count

(continued)

Table 1. (continued)

Article Reference (Study Period)	Attribute Examined	Context	Study Design	Health Condition/ Patient Population ^a	EHR Component or Data Flow	Data Elements
Benin et al., 2005 (2001–2002)	Accuracy	U.S., a large academic pediatric clinic	Comparative: admin data, EMR-based search term strategy vs. manual review ("reference")	Pediatrics; pharyngitis; children 3–18 years	Data extraction (search- term); structured fields, CPOE	Diagnosis (pharyngitis), diagnostic testing for GAS (group A streptococcus) and antibiotic prescription
Conroy et al., 2005 (2002)	Completeness	U.S., 114 providers from 8 Boston- area primary care practices affiliated with an IDS	Comparative: manual review of EHR data (structured and free- text) vs. patient and provider survey	Tobacco counseling; 522 current smokers (patient: 47%; provider: 88%; response rate)	Data documentation; structured fields and free-text	5A (tobacco cessation guidelines); smoking status (structured), date of counsel (structured)
de Lusignan et al., 2005 (2003– 2004)	Completeness; accuracy; comparability	Non-U.S. (U.K.), 78 volunteer general practices	Descriptive	Osteoporosis	Data documentation; data extraction	Age, gender, diagnoses, risk factors, osteoporosis treatment
Hazlehurst et al., 2005 (2003– 2004)	Accuracy	U.S., 4 HMOs (outpatient records)	Comparative: MediClass (NLP) vs. manual abstraction	Smoking cessation; 500 known smokers	Data extraction (NLP)	5A (tobacco cessation guidelines); smoking status assessment (ASK); counseling (ADVISE), readiness assessment (ASSESS), treatment/ support (ASSIST), follow-up (ARRANGE) in free-text
Staroselsky et al., 2006 (2003– 2004)	Accuracy	U.S., single primary care practice	Comparative: EHR data vs. patients survey	Preventive services; N = 163 (14.8% response rate)	Not specified	Flu vaccination, mammograms, pap smears, cholesterol screening, and osteoporosis screening
de Lusignan et al., 2004 (2000– 2001)	Completeness	Non-U.S. (U.K.), 87 general practices	Pre-post trial on educational intervention for improving data quality	Ischemic heart disease; N = 19,470 patients	Data documentation; not specified	IHD diagnosis, BP, aspirin prophylaxis, cholesterol measurement and control, smoking status, and smoking cessation counseling

(continued)

Table I. (continued)

Article Reference (Study Period)	Attribute Examined	Context	Study Design	Health Condition/ Patient Population ^a	EHR Component or Data Flow	Data Elements
Faulcomer and de Lusignan, 2004 (not specified)	Completeness; accuracy	Non-U.S. (U.K.), single general practice	Comparative: EHR data vs. other clinical sources (free-text note/record)	COPD; 10,975 registered GP patients	Data documentation; not specified	COPD diagnosis, smoking status, FEV1 results
Hicklen et al., 2004 (not specified)	Comparability	U.S., single IDS	Descriptive: presence, match in structure and meaning across 3 data systems (90 comparisons)	Prenatal, labor, and delivery, NICU, mother/infant	Data repositories	Not specified (30 common data elements)
Kaboli et al., 2004 (not specified)	Completeness; accuracy	U.S., primary care clinic at Iowa City VAMC	Comparative: manual EHR review vs. semistructured patient interview ("reference," meds brought to interview)	Not specified; 65 years+ with 5 or more active prescribed meds (N = 493)	Data documentation; medication list, allergy lists, adverse drug reactions	Medications (number and names), allergy (known drug sensitivity or hypersensitivity), and adverse drug reaction
Owen et al., 2004 (1998-1999)	Accuracy	U.S., 3 VA Health Care Network	Comparative: automated EHR data vs. manual review ("reference")	Schizophrenia; random sample N = 261 (outpatient and inpatient)	Data extraction	Dates, diagnosis, global assessment of functioning for encounters and stays, prescription (from pharmacy data)

Note: EHR = electronic health record; CHF = congestive heart failure; ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker; LDL = low-density lipoprotein; HDL = high-density lipoprotein; EPRP = external peer review program; CAD = coronary artery disease; ADL = activities of daily living; ART = antiretroviral therapy; CPOE = computerized physician order entry; NLP = natural language processing; COPD = chronic obstructive pulmonary disease; NICU = neonatal intensive care unit; EMR = electronic medical record; DMARD = disease modifying antirheumatic drug; HMO = health maintenance organization.

a. Some studies used different samples for different analyses.

Table 2. Data Completeness and Accuracy Findings

Study	Sample	Findings ^a by Data Element
Data completeness		
McGinnis et al. (2009)	HIV (VA patients)	HIV-specific database: hemoglobin (13%), aspartate aminotransferase (14%), alaine aminotransferase (12%), glycosylated hemoglobin (20%), creatinine (14%), white blood cell (13%), glucose (15%), HIV RNA viral load (2%), and CD4+ count (2%); All veterans in care database: hemoglobin (9%), aspartate aminotransferase (1%), alaine aminotransferase (14%), glycosylated hemoglobin (5%), creatinine (8%), white blood cell (9%), glucose (4%), HIV RNA viral load (37%), and CD4+ count (36%)
Linder et al. (2009)	Pneumonia	Smoking assessment: 36%, smoking cessation: 53%, influenza vaccination: 0%, pneumococcal vaccination: 0%, temperature: 52%, blood pressure: 51%, pulse: 58%, respiratory rate: 93%
Persell, Dunne, et al. (2009)	Adult primary care patients	For total population: blood pressure: 1.9%, total cholesterol: 22.5%, LDL: 29.1%, triglycerides: 25.3%, HDL: 23.3%. For high cardiac risk groups, missing for these outcomes range from 0.1% (blood pressure) to 4.6% (LDL, triglycerides)
Forster et al. (2008)	HIV	Age: 2.6%, gender: 0%, year ART initiated: 1.0%, HIV clinical stage: 3.5%, baseline CD4: 19.8%, follow-up CD4: 5.3% (median %)
Goodyear-Smith et al. (2008)	Children	Immunization receipt: 20% for 15 months immunization, 55% for 5 months immunization, 60% for 3 months immunization, and 70% for 6 weeks immunization
Rosenman et al. (2008)	Pregnant women	Syphilis screening reports: 1%
Goulet et al. (2007)	VA patients	LDL: 33%, blood pressure: 12%, HgA1c: 13%
Conroy et al. (2005)	Current smokers	ASK: 38%, ADVISE cessation: 73%, ASSESS readiness to quit: 77%, ASSIST cessation: 80%, ARRANGE follow-up: 98%
Faulconer and de Lusignan (2004)	COPD	FEV-1 (within 27 months): 90%; smoking status: 10%
Data accuracy		
McGinnis et al. (2009)	HIV (VA patients)	Correlation of overlapping values from two VA data repositories: hemoglobin (.997), aspartate aminotransferase (.997), alaine aminotransferase (.999), glycosylated hemoglobin (.991), creatinine (.990), white blood cell (.961), glucose (.940), HIV RNA viral load (.989), and CD4+ count (.875)

(continued)

Table 2. (continued)

Study	Sample	Findings ^a by Data Element
Persell, Dunne, et al. (2009)	Adult primary care patients	28% of 500 charts had discrepancy in age, gender, blood pressure, mean total and HDL cholesterol, medications (antihypertensive, lipid-lowering, or antithrombotic), or smoking status
Goodyear-Smith et al. (2008)	Children	Error in immunization receipt dates range from 3% to 11%
Baker et al. (2007)	Congestive heart failure	Heart failure diagnosis: 92.3% (477 of 517 patients) correctly identified through automated review of EHR data
Goulet et al. (2007)	VA patients	Correlation between structured EHR data and full chart review ($r = .89$ -BP; $r = .98$ -LDL; $r = .98$ -HbA1c)
Tang et al. (2007)	Diabetes	Diabetes diagnosis: 97.6% (sensitivity), 99.6% (specificity)
Staes et al. (2006)	Liver transplant patients	1 error found in validation of 163 laboratory results of creatinine, cyclosporin A, and tacrolimus
Staroselsky et al. (2006)	Adult primary care patients	Patient provided additions to EHR increased health maintenance adherence rates for breast cancer (84.3% to 96.3%), cervical cancer (73.2% to 96.3%), cholesterol (96.9% to 97.5%), osteoporosis (98.2% to 99.4%), and influenza vaccine (60.7% to 84.7%)
Powell et al. (2006)	Primary care patients	31 patients identified 6 verified errors in dates or clinical information
Saigh et al. (2006)	Primary care patients	55% of 97 encounters had active pain documented in free-text or the problem list, but a "no pain" entry in the data template
Faulconer and de Lusignan (2004)	COPD	COPD diagnosis: 79% (sensitivity), 75% (PPV); Providers selected an ambiguous code of "smoking status" for 5.8% ($N = 171$)
Owen et al. (2004)	Schizophrenia	Guideline concordant dosing ($\kappa = .63$; 84% agreement) and antipsychotic medications ($\kappa = .17$; 59% agreement) among outpatients

Note: LDL = low-density lipoprotein; HDL = high-density lipoprotein; ART = antiretroviral therapy; COPD = chronic obstructive pulmonary disease; EHR = electronic health record.

a. Missing data percentage presented for data completeness; various statistics presented for data accuracy.

adult primary care population. Similarly, McGinnis et al.'s (2009) study found higher levels of data completeness for viral load and CD4+ count in a database for HIV patients compared with a database for all veterans in care.

Problem lists. Whether a health condition is documented in the problem list varies by condition, with chronic conditions more likely to be documented than acute conditions. Baker et al. (2007) found 71% and Tang, Ralston, Arrigotti, Qureshi, and Graham (2007) found 93% of patients in their studies had their condition, congestive

heart failure and diabetes, respectively, documented in the problem list. In contrast, among 195 patients identified with pneumonia, only 12% had the condition documented as an active or inactive condition in the EHR's problem list (Linder, Kaleba, & Kmetik, 2009).

Medication lists. In the ambulatory setting, the omission rate for medication data was substantial across different clinical populations, from 27% of medications for ambulatory oncology patients (Weingart et al., 2007) to 53% for primary care patients (Varkey et al., 2007). Among VA patients aged ≥ 65 ($N = 493$) with 5 or more active prescribed medications, an average of 3.1 omissions per patient was observed (Kaboli et al., 2004). Prescribed medications did not fare better than other types of medications. Among 163 adult primary patients, 30% had prescribed medications and 48% had over-the-counter or herbal medications that were missing from the medication list (Staroselsky et al., 2008). Similarly, 34% of all omissions observed among study patients were for prescribed medications (Kaboli et al., 2004).

Computerized physician order entry (CPOE). There is suggestive evidence that order entry may provide better data than entered laboratory results for determining clinician performance. Benin et al. (2005) compared computerized physician order entry data with data from a structured data field designated for laboratory results for identifying diagnostic test performance. When the laboratory results field was used, 33% of patients were identified as having had the test. This compares with 71% for order entry data and 77% for the full manual EHR review. Orders data may be more complete because they are automatic by-products of clinical action. In contrast, the laboratory results field likely required manual entry. This may contribute to missing data as this information can be entered in other locations in the EHR, including free-text fields.

Data capture. Substantial variation in the amount of missing data has been observed across different clinical locations, even within the same organization. For example, missing data ranged between 24% and 38% for LDL (low-density lipoprotein), 3% and 31% for blood pressure, and 5% and 23% for HbA1c results across 8 VA clinical sites (Goulet et al., 2007). Wide site-level variation was also observed for the prevalence of diagnostic, risk factor, and treatment data across 78 U.K. general practices (de Lusignan et al., 2005).

Studies reporting that the same clinical content can be found in different or multiple locations within the EHR suggest that the variation in data completeness reflects differences in documentation practices or data capture methods used at different sites. For example, Baker et al. (2007) found congestive heart failure diagnosis recorded in numerous locations within the EHR, including encounter diagnostic fields (95%), problem list (71%), and medical history (37%) of the EHR. Another study from the Netherlands found that nearly 28% of free-text entries on patient history and physical exam could have been captured in a structured field (Roukema et al., 2006). These findings suggest that if data are not extracted from the appropriate locations in the EHR, or if data are regularly recorded in a free-text field, automatic extraction will likely produce higher missing data rates.

Data documentation rates can be improved through staff training and education. An evaluation of an educational intervention in 87 general practices in the United Kingdom found that documentation increased by 50% for smoking cessation recommendation, 10% for smoking status, 10% for cholesterol measured, and 4% for aspirin status among patients with ischemic heart disease following an intervention (de Lusignan, Hague, Brown, & Majeed, 2004). Forster et al. (2008) found that training data management staff at HIV treatment sites in resource-limited countries can also significantly reduce missing data (adjusted odds ratio = 0.58, 95% confidence interval = 0.37-0.90). In addition, they found that greater data clerk time spent on entry was associated with less missing data, with about 10 weekly hours per 100 patients needed to reduce missing data below 10%.

Workflow redesign can have beneficial effect on data quality. In a U.S. study, the documentation rate of vital signs increased from 0.7% to 58.5% of visits when standardized vital sign data fields were introduced with a redesigned workflow at three clinics (Szpunar, Williams, Dagrosa, Enberg, & Chesney, 2008). However, variation in the degree of improvement across sites (i.e., 2% to 22%, 0.5% to 58%, and 0% to 95%) continues to emphasize the influence of local practices and environment. Furthermore, not all interventions are successful. A new mandatory computerized pain assessment screen did not improve the documentation of pain because of perceived burden of using the new application (Saigh, Triola, & Link, 2006). Greater consideration of workflow issues and user needs, including training and support, during redesign should enhance effectiveness of future interventions.

Data extraction. Automated extraction of structured EHR data performed reasonably well across different types of data. Specifically, Tang et al. (2007) found that, compared with “gold standard” expert review, automated EHR data extraction identified 97% of patients with diabetes. Rosenman et al. (2008) found that laboratory reports in the EHR identified 99% of prenatal syphilis screens. Finally, Baker et al. (2007) found that automated extraction identified 94.6% of congestive heart failure patients with left ventricular ejection factor data and between 70% (warfarin) and 94% (angiotensin-converting enzyme [ACE] or angiotensin receptor blocker [ARB]) of appropriate prescriptions when compared with hybrid review (automated extraction plus manual review of electronic notes).

However, the automated extraction can also miss important data and appropriate exclusions. Manual review of free-text fields in the Baker et al. (2007) study increased data completeness by between 1.9% (beta-blocker) and 23.2% (warfarin). Goulet et al. (2007), using VA data, found lower data completeness when comparing automated EHR data abstraction with manual EHR review (LDL: 67.4% vs. 88.4%, $p < .01$; HbA1c: 87.3% vs. 93.5%, $p < .01$). Surprisingly, completeness for data elements electronically abstracted from the same EHR system can also vary. A comparative study of HIV-related data elements found that absolute differences in the percentage of documented values were between 2% (alanine aminotransferase) and 35% (viral load) in two data repositories from the VA's EHR system (McGinnis et al., 2009). These findings highlight the influence of both data extraction methods and documentation practices on data completeness.

Data Accuracy

Although data accuracy studies are also characterized by diversity in study focus, emerging evidence from these studies points to the potential of free-text data abstraction programs for accurately extracting data and the need to improve data accuracy in medication lists.

Data element content. Few studies of data accuracy examined the same data elements. However, laboratory results as a data category appear to exhibit greater data accuracy (Goulet et al., 2007; McGinnis et al., 2009; Staes et al., 2006) than other types of data. These findings may reflect the clinical importance of laboratory data and the growing use of electronic data linkage between large laboratories and health providers. Table 2 presents findings on the accuracy of a variety of diagnostic, clinical performance and patient status data from the studies we reviewed.

Medication lists. All studies of medication lists report significant errors. A high proportion of patients, 81% and 95% reported in two studies, had errors in their medication list (Kaboli et al., 2004; Weingart et al., 2007). Errors because of retention of discontinued medications were common, ranging between 13% and 29% across different patient populations (Kaboli et al., 2004; Staroselsky et al., 2008; Weingart et al., 2007). Incorrect medication regimens, less common at 5% to 14%, were also found (Staroselsky et al., 2008; Weingart et al., 2007). Prescription medications do not appear to be protected from errors as evidenced by the 66% commission error rate observed by Kaboli et al. (2004). Varkey et al. (2007) did not find most errors to have clinically important effects, classifying 24% as significant and 0.3% as serious. However, these errors may have more important effects from a quality measurement perspective.

Substantial improvement in medication list accuracy will be needed before this data source can be relied on for quality measurement. However, few effective interventions to improve medication list accuracy have been identified. Among the few published evaluations, the interventions studied have been varied, sample sizes were small, and findings have been mixed. Varkey et al. (2007) reported that a multifaceted intervention involving the care team and the patient reduced the average number of medication discrepancies per patient by more than 50% from 5.24 to 2.46. However, providing patients with access to their medication data through a Web portal did not reduce medication discrepancies (Staroselsky et al., 2008).

Data extraction. Automated abstraction of structured EHR data generally performed well, accurately producing laboratory test data and identifying medications and candidates for therapy. For example, Goulet et al. (2007) found strong agreement (κ between .86 and .99) and high sensitivity and specificity ($\geq .95$) for quality measures based on electronically abstracted structured data compared with manual review. Similarly, Baker et al. (2007) found high sensitivity ($\geq .95$) for beta-blockers, ACE or ARB, and warfarin compared with manual EHR review, and Persell, Dunne, et al. (2009) reported high sensitivity and specificity (between .81 and .98) in identifying candidates for anti-platelet therapy and lipid lowering therapy. The high level of accuracy likely reflects the clinical value of the data. Important clinical data are likely to be accurate whether

it is documented in a structured field or in a free-text field. Therefore, while electronic abstraction may miss data documented in free-text and lead to lower data completeness, electronically abstracted data can still be accurate. However, the specific rules embedded in a data extraction program can affect data accuracy. In a U.K. study, de Lusignan et al. (2005) found that the abstraction software used a more general code to extract data for hormone replacement therapy, resulting in overestimation in the number of these prescriptions.

Important clinical information is often recorded within free-text sections of the EHR. Evaluations of free-text data extraction programs, particularly natural language processing (NLP), suggest that these programs can provide accurate data on service provision (Baldwin, 2008; Hazlehurst et al., 2005; Pakhomov, Bjornsen, Hanson, & Smith, 2008) and patient clinical status (Pakhomov, Jacobsen, Chute, & Roger, 2008). For example, Pakhomov, Bjornsen, et al. (2008) found an NLP program to have overall accuracy of 75% for identifying all three components of a foot exam among diabetics. Similarly, Hazlehurst et al. (2005) found reasonably good accuracy of another NLP program (MediClass), for determining adherence to smoking cessation guidelines in current smokers. Specificity was particularly good, ranging between .82 and 1.0. Sensitivity was higher for smoking status assessment and cessation treatment ($\geq .97$) than for the other actions ($\leq .68$). Pakhomov, Jacobsen, et al. (2008) reported that an NLP program can also achieve high sensitivity for identifying the presence (.91-1.00) and absence (.83-.88) of heart disease symptoms (i.e., cough, dyspnea, and chest pain).

Finally, Baldwin's (2008) study of breast cancer prevention services found that an NLP program was able to retrieve 100% data on gender and race but demonstrated a 35% false negative rate for mammography and clinical breast exam. The authors suggested that the use of different terms to express the same concept (e.g., "CBE" or "breast exam") and the documentation location of these terms may have contributed to the observed false negatives. The NLP program appears to function more effectively for variables that are narrowly and consistently defined. These findings highlight the role of documentation practices in the accuracy of data obtained using free-text abstraction programs. For example, copying and pasting, which has been observed in 25% of charts in one study (Thielke, Hammond, & Helbig, 2007), may contribute to errors by increasing the likelihood that new information are not recorded, leading to data omissions, or old findings carried forward as new findings, leading to incorrect data. The effect of these and other documentation practices, particularly in conjunction with text-based abstraction methods, on data accuracy should be more thoroughly investigated.

Although NLP programs are more common, data abstraction using search-term methods also performed reasonably well. Benin et al. (2005) evaluated this method and found high sensitivity (96%), but low specificity (34%), in identifying cases of pharyngitis. Performance rates were similar using the search-term strategy and the reference manual EHR review, with only a 3% difference for antibiotic prescribed and no difference for Group A streptococcus test.

Data Comparability

Only a few studies explicitly examined data comparability. However, these studies have found substantial variation in data documentation, retrieval, and content across clinical sites and data systems. In particular, variation in missing data by clinical site has been reported in multiple studies. Goulet et al. (2007) found that missing data differed by 14% for LDL, 18% for HbA1c, and 28% for blood pressure values across eight VA clinical sites. Interventions, even when effective for increasing overall documentation rates, does not necessarily reduce site-level variation. Szpunar et al. (2008) found that after an intervention on data quality, structured vital sign field use rose from 0% to 2% across the three study sites to 22%, 58%, and 95.2%. Site variation has also been reported for the recording of childhood immunization receipt and dates across 124 general practices (Goodyear-Smith et al., 2008). de Lusignan et al. (2005) found that different codes can be used to label patients of similar clinical status, leading to data unreliability across the 78 practices in their study. Furthermore, systemic differences in missing data by practice were common; with some practices missing data on steroids, others lacking data on fractures or medications. Data extraction problems appeared to have contributed to these problems and those observed in McGinnis et al. (2009) where different rates of missing data were found for the same elements in two data repositories, even though the repositories draw data from the same EHR.

Ricciardi, Lieberman, Kahn, and Masarie (2005) reported on terminology harmonization efforts on problem list entries with a consortium of 79 organizations using the same commercial ambulatory EHR (GE Centricity). Despite evidence of successful mapping (86% of entries), the study also highlighted substantial variability in how users document in the problem list. Even within the same organization, data comparability is not assured. Hicken, Thornton, and Rocha (2004) found that of 30 elements across three legacy systems spanning prenatal, labor and delivery, and mother/infant care in the same organization, only 5 elements (17%) had the same meaning and structure (format, granularity). Another 41 elements (46% of 90 comparisons) were related in meaning but had different granularity across systems. Local naming conventions (McGinnis et al., 2009) and locally developed glossaries for selected coded information such as problems, vaccinations, and medications (Linder et al., 2009) also appears to contribute to the lack of data comparability across clinical sites.

Overall, these findings suggest that poor data comparability will be an important impediment to the use of EHR data for quality reporting. To enhance data comparability, health care organizations will need to make careful decisions regarding the data source used for quality measurement, better standardize data documentation and extraction procedures, and implement effective interventions or tools to facilitate data harmonization across organizations.

Measure Validity

The wide array of readily extractible treatment and clinical elements from the EHR can produce more refined and valid measures of care quality. Persell, Kho, et al. (2009) used structured EHR data to craft hybrid process and outcome quality measures that better capture the quality of hypertension care. For hypertensive patients without diabetes, incorporating relevant clinical data to refine inclusion and exclusion criteria can improve measure validity. For example, measures can now account for aggressive prescribed treatment or low diastolic BP, which raised performance from 75% to 83% and 84%, respectively. Including patients with undiagnosed hypertension (i.e., mean systolic BP ≥ 140 or diastolic BP ≥ 90 over their last three visits) to the denominator lowered performance to 81%.

However, documentation practices and EHR design, particularly the use of free-text fields, can compromise measure validity by making it difficult to correctly identify eligible cases or appropriate exclusions. Persell et al. (2006) examined cases of quality failures for seven coronary heart disease measures, which had been automatically calculated using structured EHR data. Through manual review of free-text notes, the authors found that 15% to 81% of cases identified as quality failures actually satisfied the performance measure or met appropriate exclusion criteria. Baker et al. (2007) confirmed that appropriate exclusions are often missed when an automated extraction method based on data from structured fields is used to calculate a measure, reporting sensitivity ranging between 49% (warfarin) and 72% (ACE or ARB). However, it is important to note that these studies were limited by the lack of a true gold standard, making measure validity difficult to determine.

Discussion

This article provides an overview of the current state of the literature on the quality of EHR data from the perspective of quality measurement. Although we found a growing number of research studies, data attributes that are important for quality measurement, including timeliness, granularity, and comparability, have not been adequately studied. Furthermore, the diversity in study focus and the relatively weak study designs used in the current body of research made it difficult to draw clear conclusions about the data quality of specific data elements or EHR components.

However, preliminary observations regarding EHR data quality can be made. First, data reliability from the problem and the medication lists remains questionable. Significant errors of omission and commission were found for medication list and data quality appears to vary by condition in the problem list. Chronic diseases documented in the problem list appear more accurate than for other acute conditions, although chronically ill patients may still be missed. Given these findings, the data quality from these sources must improve significantly before they can be used for quality reporting.

Second, local data documentation practices and extraction methods exert strong effects on data quality, particularly data completeness. Clinical sites, even within the

same organization and using the same EHR, may rely on electronically linked data sources such as commercial laboratories to varying degrees, particularly if their patients have different health plans or are geographically dispersed (Staes et al., 2006). Even among VA patients, outside services including laboratory tests may be used, particularly if the patient is also covered by other insurance. For nonelectronically linked data, clinical staff often has multiple options within the EHR, from structured fields to the free-text notes, for documentation. This may account for the higher data completeness observed with manual rather than electronic abstraction, as automated abstraction usually relies on structured fields and miss data in free-text fields that manual review can identify. Given these site- and staff-level variability, data comparability for the same commercial EHR product or even within a single organization should not be assumed. However, there is promising evidence that careful design and implementation of new data documentation and abstraction procedures can minimize variability in data completeness and significantly enhance the value of EHR data for quality measurement.

Finally, the reliability of free-text data extraction programs, such as NLP, appears good, particularly for narrowly and consistently defined variables. However, the specific data extraction strategy used and the documentation style of users in the study can affect findings, which limits the generalizability of these studies.

Directions for Future Research

One of the key insights that emerged from this review is the need for a stronger evidence base to support the use of EHR data in quality measurement. Several research opportunities can be readily identified. First, data elements that are used for quality measurement should be identified and prioritized for empirical investigation. These studies should include data granularity and timeliness as well as data completeness and accuracy. Identifying factors that affect and methods that enhance data comparability across organizations and vendor products (e.g., granularity in menu options and data dictionaries) should be another major area of research. EHR systems vary in their interface design and their functional capabilities. In addition, these systems are often modified locally to adapt to the specific needs of each organization, further limiting data comparability even when the same commercial EHR product is used. Finally, differences in EHR design can affect provider interactions with the system, which can exacerbate differences in data reliability and availability across organizations. Additional research on these data attributes will be crucial for identifying the data source(s) that maximize data quality and comparability for high-priority data elements.

EHRs have created a wide range of new data sources, including order entry and decision support modules, audit trails, and linkages with external data sources. These new data sources offer opportunities for data supplementation and validation that should be empirically evaluated. For example, electronic data linkage with external sources such as clinical trials, local providers, social security administration, and state or local immunization registries can provide data updates on laboratory test results, service provision, and deaths. Furthermore, audit trails can provide detailed electronic tracking

of provider interactions with the EHR, such as medication reconciliation or referrals. Order entry modules provide a wide range of physician orders that can be electronically tracked. These new types of data can be used to develop measures of greater clinical validity as they can offer accurate, detailed data on relevant actions. However, apart from the problem list and medication list, few studies in our review included these EHR data sources in their assessment.

Although variation in data quality have been identified in a number of studies, more detailed examination of the factors and processes associated with the observed variation is needed. For example, documentation rates have been shown to be poor if data entry requires significant navigation (Baron, 2007; Roukema et al., 2006). In particular, the availability of free-text fields at a high level may contribute to loss of structured data if the structured data field is accessible only through multiple "clicks" or key-strokes. It has also been suggested that the clinical relevance of the data and their anticipated future use (e.g., for decision support) may affect data quality (Roukema et al., 2006; Tang et al., 2007). Standardization of data documentation and management procedures (Hicken et al., 2004) and terminologies (Linder et al., 2009) should be examined for their effectiveness in decreasing variation in documentation rates and enhancing data comparability. These and other questions are important to investigate in identifying optimal strategies for improving data quality of the EHR for quality measurement.

Interventions to improve data quality should be developed and evaluated. This may involve pilot or demonstration studies to embed data documentation supports and data quality checks in the EHR as well as workflow redesign and staff education or training. For example, opportunities to code reasons for patient's exclusion from a standard care protocol could be embedded within a clinical decision support system.

Finally, more studies, similar to those by Tang et al. (2007) and Persell, Kho, et al. (2009), should be performed to gain a better understanding how EHR data may be used to craft more refined measure specification and how to further enhance the validity of quality measures.

Looking Ahead: The Role of Provider Organizations and Third Party Organizations

This review identified important impediments, in terms of data quality and comparability, to the use of EHR data for reliable and valid quality measurement. Although additional research is clearly needed, improvements to data quality will likely be driven by innovations within health care organizations and by regulatory or competitive demands from third party entities, including accreditation organizations, governmental agencies, and payers. Each will have important roles to play in improving the quality and comparability of EHR data for quality measurement.

For health care organizations, greater attention must be paid to how clinical workflow and EHR system design interact to affect data quality. For example, the placement and design (e.g., clinical specificity and number of option choices) of structured data fields should be optimized, particularly in relation to free-text

documentation alternatives, to ensure availability of high-quality structured data. Effective data quality checks should be embedded within EHR systems (Forster et al., 2008). Furthermore, new data sharing or exchange agreements between providers and other organizations should be developed to support accurate determination of eligible cases (e.g., by removing deceased individuals based on monthly data from the Social Security Administration) and improve the completeness and timeliness of data on test results, service utilization, medications, and health conditions. Of course, data security safeguards including secure data exchange and identity verification will be required to support these changes. New EHR-based algorithms or methodologies for identifying cases should be empirically evaluated for validity. If known changes within the organization have taken place, such as addition of new clinical sites to a health system, the measured results should be checked for evidence of these events.

Regulatory bodies and accrediting organizations have particularly important roles to play in setting standards for and auditing the quality and comparability of data used for quality measurement. First, these entities should advocate for, prioritize, and specify data elements that must be reliably collected across organizations so they can be used for report cards and other quality measurement applications. They can also support and disseminate findings from independently conducted research on the reliability of different data sources and encourage diffusion of effective data collection strategies across organizations. The “e-provenance” of each electronically derived quality measure could be requested from health care organizations. These may include the frequency with which key data elements are updated and their process for capturing, storing, and extracting the quality indicator. For example, are comparable data source(s) used for required data elements? Are common nomenclatures or coding schemes used? If not, do they differ substantially in granularity? Do the data capture and extraction methods, including automated structured data and free-text abstraction programs, used by different health care organizations produce data of comparable reliability? What internal validity checks for entered data are performed? Are there plans for regular monitoring of data quality and procedures for resolving data reliability problems? As appropriate, information gleaned from these investigations for different organizations could be used to benchmark each organization’s progress, and a process to share best practices should be developed. Finally, standards for the reliability, granularity, and format of high-priority data elements should be clearly articulated to further enhance data comparability.

Limitations

We have focused our review on peer-reviewed articles. Many of these articles described studies conducted in large, often academically affiliated, health care organizations. The experiences of small and solo practices with data reliability and quality measurement are not well represented, thus our review likely represents the best case. In addition, although our search strategy has identified a rich and varied set of studies for review, it is possible that some articles relevant to our review may be missed because of the

wide range of keywords under which articles may be indexed. Despite these limitations, we believe our review provides a generally accurate depiction of the state of research in this area and permit us to identify major directions for future research.

Conclusion

Issues related to data accuracy, completeness, and comparability must be addressed before routine EHR-based quality of care measurement can be done with confidence. We expect that the range and type of data that can be reliably and comparably collected through EHRs will increase as system capacity grows, users become more comfortable with new health IT systems, and continuous quality improvement principles are applied to the EHR-based data collection–storage–retrieval process. Furthermore, the accuracy and value of EHR-derived data will improve as evidence grows on the relative merits of different EHR data sources and procedures for implementing HIT-supported quality measures.

Declaration of Conflicting Interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: This project is co-funded by The Commonwealth Fund (#20060073), The Robert Wood Johnson Foundation (#053729), and the Agency for Healthcare Research and Quality (#6275-JHU-01).

References

- Agnew-Blais, J. C., Coblyn, J. S., Katz, J. N., Anderson, R. J., Mehta, J., & Solomon, D. H. (2009). Measuring quality of care for rheumatic diseases using an electronic medical record. *Annals of the Rheumatic Diseases*, 68, 680-684.
- Baldwin, K. B. (2008). Evaluating healthcare quality using natural language processing. *Journal for Healthcare Quality*, 30, 24-29.
- Baker, D. W., Persell, S. D., Thompson, J. A., Soman, N. S., Burgner, K. M., Liss, D., et al. (2007). Automated review of electronic health records to assess quality of care for outpatients with heart failure. *Annals of Internal Medicine*, 146, 270-277.
- Baron, R. J. (2007). Quality improvement with an electronic health record: Achievable, but not automatic. *Annals of Internal Medicine*, 147, 549-552.
- Benin, A. L., Vitkauskas, G., Thornquist, E., Shapiro, E. D., Concato, J., Aslan, M., et al. (2005). Validity of using an electronic medical record for assessing quality of care in an outpatient setting. *Medical Care*, 43, 691-698.
- Conroy, M. B., Majchrzak, N. E., Silverman, C. B., Chang, Y., Regan, S., Schneider, L. I., et al. (2005). Measuring provider adherence to tobacco treatment guidelines: A comparison of electronic medical record review, patient survey, and provider survey. *Nicotine and Tobacco Research*, 7(Suppl. 1), S35-S43.

- Dean, B. B., Lam, J., Natoli, J. L., Butler, Q., Aguilar, D., & Nordyke, R. J. (2009). Use of electronic medical records for health outcomes research: A literature review. *Medical Care Research and Review*, 66, 611-638.
- de Lusignan, S., Chan, T., Wood, O., Hague, N., Valentin, T., & Van Vlymen, J. (2005). Quality and variability of osteoporosis data in general practice computer records: Implications for disease registers. *Public Health*, 119, 771-780.
- de Lusignan, S., Hague, N., Brown, A., & Majeed, A. (2004). An educational intervention to improve data recording in the management of ischaemic heart disease in primary care. *Journal of Public Health*, 26, 34-37.
- desRoches, C. M., Campbell, E. G., Rao, S. R., Donelan, K., Ferris, T. G., Jha, A., et al. (2008). Electronic health records in ambulatory care: A national survey of physicians. *New England Journal of Medicine*, 359, 50-60.
- Faulconer, E. R., & de Lusignan, S. (2004). An eight-step method for assessing diagnostic data quality in practice: Chronic obstructive pulmonary disease as an exemplar. *Informatics in Primary Care*, 12, 243-253.
- Fishman, P. A., Hornbrook, M. C., Meenan, R. T., & Goodman, M. J. (2004). Opportunities and challenges for measuring cost, quality and clinical effectiveness in health care. *Medical Care Research and Review*, 61, 124S-143S.
- Forster, M., Bailey, C., Brinkhof, M. W., Graber, C., Boulle, A., Spohr, M., et al. (2008). Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings. *Bulletin of the World Health Organization*, 86, 939-947.
- Goodyear-Smith, F., Grant, C., York, D., Kenealy, T., Copp, J., Petousis-Harris, H., et al. (2008). Determining immunisation coverage rates in primary health care practices: a simple goal but a complex task. *International Journal of Medical Informatics*, 77, 477-485.
- Goulet, J. L., Erdos, J., Kancir, S., Levin, F. L., Wright, S. M., Daniels, S. M., et al. (2007). Measuring performance directly using the Veterans health administration electronic medical record: A comparison with external peer review. *Medical Care*, 45, 73-79.
- Hazlehurst, B., Sittig, D. F., Stevens, V. J., Smith, K. S., Hollis, J. F., Vogt, T. M., et al. (2005). Natural language processing in the electronic medical record: Assessing clinician adherence to tobacco treatment guidelines. *American Journal of Preventive Medicine*, 29, 434-439.
- Hicken, V. N., Thornton, S. N., & Rocha, R. A. (2004). Integration challenges of clinical information systems developed without a shared data dictionary. *Studies in Health Technology and Informatics*, 107, 1053-1057.
- Institute of Medicine. (2006). *Rewarding provider performance: Aligning incentives in Medicare*. Washington, DC: National Academies Press.
- Kaboli, P. J., McClimon, B. J., Hoth, A. B., & Barnett, M. J. (2004). Assessing the accuracy of computerized medication histories. *American Journal of Managed Care*, 10, 872-877.
- Linder, J. A., Kaleba, E. O., & Kmetik, K. S. (2009). Using electronic health records to measure physician performance for acute conditions in primary care: Empirical evaluation of the community-acquired pneumonia clinical quality measure set. *Medical Care*, 47, 208-216.
- McGinnis, K. A., Skanderson, M., Levin, F. L., Brandt, C., Erdos, J., & Justice, A. C. (2009). Comparison of two VA laboratory data repositories indicates that missing data vary despite originating from the same source. *Medical Care*, 47, 121-124.

- Owen, R. R., Thrush, C. R., Cannon, D., Sloan, K. L., Curran, G., Hudson, T., et al. (2004). Use of electronic medical record data for quality improvement in schizophrenia treatment. *Journal of the American Medical Informatics Association*, 11, 351-357.
- Pakhomov, S., Bjornsen, S., Hanson, P., & Smith, S. (2008). Quality performance measurement using the text of electronic medical records. *Medical Decision Making*, 28, 462-470.
- Pakhomov, S. V., Jacobsen, S. J., Chute, C. G., & Roger, V. L. (2008). Agreement between patient-reported symptoms and their documentation in the medical record. *American Journal of Managed Care*, 14, 530-539.
- Persell, S. D., Dunne, A. P., Lloyd-Jones, D. M., & Baker, D. W. (2009). Electronic health record-based cardiac risk assessment and identification of unmet preventive needs. *Medical Care*, 47, 418-424.
- Persell, S. D., Kho, A. N., Thompson, J. A., & Baker, D. W. (2009). Improving hypertension quality measurement using electronic health records. *Medical Care*, 47, 388-394.
- Persell, S. D., Wright, J. M., Thompson, J. A., Kmetik, K. S., & Baker, D. W. (2006). Assessing the validity of national quality measures for coronary artery disease using an electronic health record. *Archives of Internal Medicine*, 166, 2272-2277.
- Powell, J., Fitton, R., & Fitton, C. (2006). Sharing electronic health records: The patient view. *Informatics in Primary Care*, 14, 55-57.
- Pringle, M., Ward, P., & Chilvers, C. (1995). Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. *British Journal of General Practice*, 45, 537-541.
- Ricciardi, T. N., Lieberman, M. I., Kahn, M. G., & Masarie, F. E., Jr. (2005). Clinical terminology support for a national ambulatory practice outcomes research network. *AMIA Annual Symposium Proceedings*, 2005, 629-633.
- Rosenman, M. B., Tao, G., Szucs, K. A., Wang, J., Ambuehl, R., & Mahon, B. E. (2008). Prenatal syphilis screening rates measured using Medicaid claims and electronic medical records. *Sexually Transmitted Diseases*, 35, 387-392.
- Roukema, J., Los, R. K., Bleeker, S. E., van Ginneken, A. M., van der Lei, J., & Moll, H. A. (2006). Paper versus computer: Feasibility of an electronic medical record in general pediatrics. *Pediatrics*, 117, 15-21.
- Saigh, O., Triola, M. M., & Link, R. N. (2006). Brief report: Failure of an electronic medical record tool to improve pain assessment documentation. *Journal of General Internal Medicine*, 21, 185-188.
- Schneider, E. C., Riehl, V., Courte-Wienecke, S., Eddy, D. M., & Sennett, C. (1999). Enhancing performance measurement: NCQA's road map for a health information framework. *Journal of the American Medical Association*, 282, 1184-1190.
- Staes, C. J., Bennett, S. T., Evans, R. S., Narus, S. P., Huff, S. M., & Sorensen, J. B. (2006). A case for manual entry of structured, coded laboratory data from multiple sources into an ambulatory electronic health record. *Journal of the American Medical Informatics Association*, 13, 12-15.
- Staroselsky, M., Volk, L. A., Tsurikova, R., Newmark, L. P., Lippincott, M., Litvak, I., et al. (2008). An effort to improve electronic health record medication list accuracy between

- visits: Patients' and physicians' response. *International Journal of Medical Informatics*, 77, 153-160.
- Staroselsky, M., Volk, L. A., Tsurikova, R., Pizziferri, L., Lippincott, M., Wald, J., et al. (2006). Improving electronic health record (EHR) accuracy and increasing compliance with health maintenance clinical guidelines through patient access and input. *International Journal of Medical Informatics*, 75, 693-700.
- Szpunar, S., Williams, P., Dagroso, D., Enberg, R., & Chesney, J. (2008). Improving the documentation of vital signs: A business reengineering efficiency study. *Joint Commission Journal on Quality and Patient Safety*, 34, 171-178.
- Tang, P. C., Ralston, M., Arrigotti, M. F., Qureshi, L., & Graham, J. (2007). Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: Implications for performance measures. *Journal of the American Medical Informatics Association*, 14, 10-15.
- Thielke, S., Hammond, K., & Helbig, S. (2007). Copying and pasting of examinations within the electronic medical record. *International Journal of Medical Informatics*, 76(Suppl. 1), S122-S128.
- Thiru, K., Hassey, A., & Sullivan, F. (2003). Systematic review of scope and quality of electronic patient record data in primary care. *British Medical Journal*, 326, 1070-1074.
- Varkey, P., Cunningham, J., & Bisping, D. S. (2007). Improving medication reconciliation in the outpatient setting. *Joint Commission Journal on Quality and Patient Safety*, 33, 286-292.
- Weingart, S. N., Cleary, A., Seger, A., Eng, T. K., Saadeh, M., Gross, A., et al. (2007). Medication reconciliation in ambulatory oncology. *Joint Commission Journal on Quality and Patient Safety*, 33, 750-757.
- Williams, J. G. (2003). Measuring the completeness and currency of codified clinical information. *Methods of Information in Medicine*, 42, 482-488.