



PERFORMANCE MEASURES USING ELECTRONIC HEALTH RECORDS: FIVE CASE STUDIES

Jinnet Briggs Fowles, Elizabeth A. Kind, and Shadi Awwad
Park Nicollet Institute

Jonathan P. Weiner and Kitty S. Chan
Johns Hopkins Bloomberg School of Public Health

Patricia J. Coon, Billings Clinic; James T. Krizak and Lynne Dancha, HealthPartners;
Nancy Jarvis, Park Nicollet Health Services; Dean F. Sittig and Brian L. Hazlehurst,
Kaiser Permanente of the Northwest; and Mark J. Selna, Geisinger Health System

May 2008

ABSTRACT: This report examines the experiences of five provider organizations in developing, testing, and implementing quality-of-care indicators, based on data collected from their electronic health record (EHR) systems. HealthPartners used the EHR to compile blood pressure measurements, Park Nicollet Health Services developed a composite measure for care of people with diabetes, Billings Clinic tested an automatic alert on potential interactions between antibiotics and the anticoagulant warfarin, Kaiser Permanente used a natural-language processing tool for counseling about tobacco use, and Geisinger Health System explored ways of reconciling Problem Lists and provider-visit notes regarding high-impact chronic-disease diagnoses. Common themes emerged from these case studies. They included challenges—of ensuring the validity and reliability of data, efficient workflow, and staff support—but the providers’ successes in implementing their respective EHR-based quality measures demonstrated that such measures are adaptable to different EHR systems, amenable to improvement, and worth pursuing.

Support for this research was provided by The Commonwealth Fund and The Robert Wood Johnson Foundation, with additional support from the Agency for Healthcare Research and Quality. The views presented here are those of the authors and do not necessarily reflect those of the two sponsors or their directors, officers, or staff. This document and other Commonwealth Fund publications are available online at www.commonwealthfund.org. To learn more about new publications when they become available, visit the Fund’s Web site and [register to receive e-mail alerts](#). Commonwealth Fund pub. no. 1132.

CONTENTS

List of Figures and Tables.....	iv
About the Authors.....	v
Executive Summary	vi
Introduction.....	1
Case Study #1 Using the EHR to Compile Blood Pressure Measurements	6
Case Study #2 Optimal Care for Diabetes: Linking Measurement with Improvement	12
Case Study #3 Warfarin/Antibiotic Rule for the EHR.....	19
Case Study #4 Assessing Clinician Adherence to Smoking-Cessation Guidelines Using MediClass: An Innovative Natural-Language Processing Tool.....	24
Case Study #5 Problem List Diagnosis Reconciliation	29
Five Case Studies: Conclusions	34
References.....	36

LIST OF FIGURES AND TABLES

Figure 1	HealthPartners EMR Data in Contrast with HEDIS Controlling High Blood Pressure Results	10
Table 1	Results for Electronic Health Records (EHRs) and Blood Pressure (BP) Readings.....	11
Figure 2	Extended Laboratory Menu (ELM) vs. NON-ELM Percent Patients Grand Slam.....	16
Table 2	The “5As” Recommended by the Current U.S. Public Health Service Clinical Practice Guideline for Tobacco Treatment and Prevention	25
Table 3	Diagnosis Specificity for Selected Chronic Diseases— Problem List vs. Encounters	33

ABOUT THE AUTHORS

Jinnet Briggs Fowles, Ph.D., is a senior vice president at the Park Nicollet Institute, where she has worked since 1985. Her areas of expertise include quality measurement, data reliability, and consumer and provider responses to information about quality. She earned her M.S. in Library Science from Simmons College and her Ph.D. from Stanford University in Communications Research.

Jonathan P. Weiner, Dr.P.H., is a professor of health policy and management at the Johns Hopkins Bloomberg School of Public Health. He is also deputy director of the Johns Hopkins Health Services R&D Center. His areas of expertise include primary care, quality of care, risk adjustment and the application of electronic health records, claims, and other computerized data to health services research and evaluation. He holds a doctorate in health services research from Johns Hopkins.

Kitty S. Chan, Ph.D., is assistant professor in the Department of Health Policy and Management at the Johns Hopkins Bloomberg School of Public Health. Her areas of expertise include psychometrics and quality measurement. She received her Ph.D. in Health Services Research from Johns Hopkins in 2001. Prior to joining the faculty at Johns Hopkins in 2004, she was associate policy researcher at the RAND Corporation between 2001 and 2004.

Elizabeth A. Kind M.S., R.N., is director of survey research at the Park Nicollet Institute's Health Research Center where she has worked since 1983. Her responsibilities include project management and overseeing instrument design and field data collection for the Center's research. She earned her master of science degree from the University of Minnesota.

Shadi Awwad, M.S., at the time of this study was a graduate assistant at the Park Nicollet Institute, completing the University of Minnesota's master's program in Health Services Research and Policy. Mr. Awwad is currently assistant to the Secretary General, Royal Hashemite Court, Jordan, Asia.

EXECUTIVE SUMMARY

The emergence of the electronic health record (EHR), also termed the electronic medical record, has made new indicators of quality and safety both necessary and feasible. By developing appropriate indicators now, we can integrate them into evolving EHR systems early on rather than try to add them after the fact—a much more difficult task. This report examines the experiences of five provider organizations in developing, testing, and implementing such indicators, based on data collected from their EHR systems.

To set the stage, we developed a typology for categorizing electronic measures (“e-indicators”) of quality and safety, with special reference to ambulatory care. The five categories are:

1. **Translational e-indicators** are measures that have been translated from existing—“traditional”—measurement sets (e.g., HEDIS or NQF standard measures) for use in health information technology (HIT) platforms.
2. **HIT-facilitated e-indicators** are measures that, while not conceptually limited to HIT-derived data sources, would not be operationally feasible in settings without HIT platforms. Measuring clinical physiologic outcomes on 100 percent of patients, for instance, would not be amenable to traditional systems.
3. **HIT-enabled e-indicators** are innovative measures that would not generally be possible outside of the HIT context. These indicators are linked to unique HIT capabilities such as computerized provider order entry (CPOE), clinical decision support systems (CDSS), biometric devices, or Web-based patient portals.
4. **HIT-system-management e-indicators** are measures needed to implement, manage, evaluate, and generally improve HIT systems. They are primarily intended for use by the parent organization.
5. **“E-iatrogenesis” e-indicators** are measures of patient harm caused at least in part by the application of health information technology. They assess the degree to which unanticipated quality and safety problems arise, whether of human (provider or patient), technical, or organizational/system origin.

The case studies presented in this report illustrate the use of the first four categories of e-indicators. The **HealthPartners** case study analyzed the potential of EHRs to compute traditional quality measures (in this case, blood pressure control) aimed at reducing the time and cost required to assemble the data. The **Park Nicollet Health Services** case

study illustrated the power of the EHR to assemble composite measures (in this case, diabetes) that are theoretically possible without an EHR but infeasible in practice. The **Billings Clinic** case study exemplified the HIT strengths of EHRs to coordinate care and measure its outcomes, in this case for a warfarin/antibiotic alert tied to a warfarin clinic. The **Kaiser Permanente of the Northwest** case study overcame the “free-text dilemma”—that free text, or unstructured information, cannot be readily used for quantitative analyses—by using natural language processing to capture information in text notes. Work at **Geisinger Health System**, meanwhile, focused on reconciling information on the health problem list (a structured-text field) with information in the visit note (an unstructured-text field).

From these case studies, a number of common themes emerged:

- It is striking how much more clinically relevant measures can become when they are HIT-based. For example, a composite measure reflects a more complete clinical picture of a person with diabetes than a single component of the measure can.
- A major barrier in conceptualizing and developing the e-indicators was the validity of EHR-extracted data, which critically depends on use of the correct patient population. If patients are incorrectly included in or excluded from a measure, the quality measures will be inaccurate.
- Another major barrier was the sometimes questionable reliability of EHR-extracted data, particularly when their collection and recording were inconsistent; the case studies suggested that it was difficult to consistently code data about patients, diagnoses, and procedures. But in addition to identifying these accuracy concerns, most of the case studies implemented workable solutions.
- Prior to implementation of the measures, providers expressed concern that EHRs would hinder workflow or suffer from staff resistance. Surprisingly, these issues did not present themselves. To the contrary, the case studies indicated that EHR systems enhanced workflow by automating key communications between staff and improving patient-record accessibility across different clinics.
- Measures that translated established quality indicators had the easiest transition into EHR implementation. Measures incorporating or evaluating HIT-specific features, such as automated alerts and free-text analysis, tended to be specialized to particular systems and not so easily incorporated into other systems. Nonetheless, most of the providers were confident that the concepts could be adapted to different EHR system types and that virtually all of the problems

encountered were amenable to performance improvement—often made possible by the EHR.

The success of these providers in implementing EHR-based quality measures demonstrates that such measures are worth pursuing, despite the challenges of ensuring the validity and reliability of data, efficient workflow, and staff support.

**PERFORMANCE MEASURES
USING ELECTRONIC HEALTH RECORDS:
FIVE CASE STUDIES**

INTRODUCTION

Electronic health records (EHRs), also termed electronic medical records, are not yet widespread in the United States, though several leading provider organizations have begun to deploy them, and they are widely considered the wave of the future. EHRs and related health information technology (HIT)—such as computerized physician order entry (CPOE), clinical decision support systems (CDSS), and Web-based patient portals—significantly enhance our ability to evaluate the processes and outcomes of health care and the degree to which consumer needs are being met. But these and other traits of EHRs call for new indicators of quality and safety. By developing appropriate indicators now, we can integrate them into evolving EHR systems early on rather than try to add them after the fact—a much more difficult task.

This report examines the experiences of five provider organizations in developing, testing, and implementing quality-of-care indicators, based on data collected from their EHR systems. While the focus of each case study was unique, they presented common strengths and weaknesses.

EHR systems should have the following features:

- *Accuracy* (validity). Data derived from patient records, both inside and outside the system, must be correct and complete.
- *Standardization* (reliability). Information taken from each patient must be standard and consistent, both across the subject population and within individual patient records.
- *Generalizability*. Measures should be able to be translated effectively between different data-collection systems.
- *Workflow efficiency*. Data collection should be structured as part of the work process to avoid negatively affecting patient visits—and ideally, to make visits more efficient.

- *Staff support.* Management and staff should together develop an infrastructure that encourages support of and adherence to the quality-measurement program.

These characteristics embrace the continuum of issues encountered—technical issues (such as validity, reliability, generalizability), human-technology interface issues (workflow efficiency), and cultural issues (staff support)—when using EHRs to support quality measurement.

Quality indicators have traditionally been developed by organizations such as the National Committee on Quality Assurance and the American Medical Association’s Physician Consortium for Performance Improvement. Developed measures have been approved for use by organizations such as the National Quality Forum and the Joint Commission for the Accreditation of Healthcare Organizations. Still other organizations, such as the Ambulatory Quality Alliance, Bridges to Excellence, and Medicare, are primarily responsible for implementing the measures.

These indicators traditionally depend on information derived from insurance claims (or other billing data), medical records, and quality-of-care surveys of patients. As many leading provider organizations have implemented large EHR systems, quality information has also increasingly come from the EHRs themselves.

EHRs and related HIT components offer many opportunities for organizations wishing to dramatically expand their quality- and safety-improvement activities or other types of performance monitoring. However, because the deployment of EHRs can also have some negative consequences, a number of related quality challenges are beginning to emerge. Often, they arise during the transition from manual to electronic systems (Ash 2004, Bates 2001, Berger 2004). But even after a system has become functional, problems can persist (Koppel 2005).

We have developed a typology for categorizing the broad types of electronic indicators (“e-indicators”) of quality and safety, with special reference to ambulatory care. It was based on our understanding of how interoperable EHRs and other types of HIT have been used to date, as well as on the potential that we, the members of our provider consortium, and the early adopters we interviewed envision they may have before long.

This typology is not meant to replace other ways of thinking about quality—for example, the Institute of Medicine’s six attributes of quality (safe, effective, patient-

centered, timely, efficient, and equitable) or Donabedian’s triumvirate of structure, process, and outcome. A new typology of e-indicators is necessary, given the functions and capabilities not seen before and the system-induced problems that have emerged, and we believe the typology can be used as an adjunct. In that way, it can help us communicate with one another as HIT-based quality improvement and monitoring systems are developed, implemented, and evaluated.

The five-class typology of electronic indicators we propose is as follows:

1. **Translational e-indicators** are measures that have been translated from existing—“traditional”—measurement sets (e.g., HEDIS or NQF standard measures) for use in HIT platforms.

Examples of translational measures include, for example, the number of patients with diabetes having an eye-care referral or the number of children receiving appropriate immunizations. Issues that surround the comparability of such translational measures have been the subject of a number of recent papers (Tang 2007).

If a traditional paper-chart-derived measure (e.g., blood-pressure readings or actual laboratory values) is expanded from a limited subsample of patients to a full patient population, it should be considered an “HIT-facilitated” measure (see below) rather than a translational measure.

2. **HIT-facilitated e-indicators** are measures that, while not conceptually limited to HIT-derived data sources, would not be operationally feasible in settings without HIT platforms.

Examples of HIT-facilitated measure include: clinical outcomes of 100 percent of patients based on physiologic measures such as Body Mass Index, blood pressure, or laboratory values; percentage of cases in which clinicians or patients receive reminders for preventive screening; and percentage of newly written prescriptions that are filled by the patient within seven days.

3. **HIT-enabled e-indicators** are innovative measures that would not generally be possible outside of the HIT context. These indicators, linked to unique HIT capabilities such as CPOE, CDSS, biometric devices, or Web-based patient portals, may involve only one HIT component or require the interaction of several.

Examples of HIT-enabled measures include: percentage of patients for whom real-time CDSS modules have been appropriately applied in support of the diagnostic process; percentage of congestive-heart-failure patients with daily e-monitored weight gain greater than x pounds, acted upon by the responsible

clinician within y hours; percentage of patients who respond appropriately to messages regarding abnormal test results; percentage of generalists who view the medical notes added by a consulting specialist within seven days of the consult; and percentage of overweight adults referred to nutrition class through a CPOE.

4. **HIT-system-management e-indicators** are measures needed to implement, manage, evaluate, and generally improve HIT systems, and they are primarily intended for use by the parent organization. However, these measures can also be used by an external body (e.g., a payer) to evaluate the organization and the HIT system it deploys for the consumer population of interest.

Examples of HIT-system-management measures include: EHR item-completion rates; attainment of community interoperability targets; a CDSS algorithm's accuracy; percentage of real-time alerts bypassed by the clinician; percentage of patient-allergy lists reviewed by patients (via Web portal) annually; and ease of access of measures that are maintained in the free-text section of the EHR.

5. **“E-iatrogenesis” e-indicators** are measures of patient harm caused at least in part by the application of health information technology. They assess the degree to which unanticipated quality and safety problems arise, whether of human (provider or patient), technological, or organizational/system origin, and they may involve errors of commission or omission (Ash 2004, Campbell 2006, Weiner 2007).

Examples of e-iatrogenesis measures include: percentage of patients receiving incorrect medications or procedures because of HIT-related errors in the CPOE process; percentage of patients experiencing a degree of harm from an unanticipated care-delivery event; percentage of significant CDSS errors (either type 1 or type 2) for a particular condition; human/machine interaction errors that lead to an incorrectly entered diagnosis; and the number of patients experiencing harm because they received another patient's orders by CPOE.

This measure is designated separately because of its key importance to those concerned with quality and safety in the HIT context.

The case studies of this report illustrate the use of the first four categories of e-indicators: translational (HealthPartners), HIT-facilitated (Park Nicollet), HIT-enabled (Billings Clinic), and HIT-system-management (Geisinger, Kaiser-Portland).

- The HealthPartners case study analyzed the potential of EHRs to compute traditional quality measures (in this case, blood pressure control) aimed at reducing the time needed to assemble the data.
- The Park Nicollet case study illustrated the power of EHRs to assemble composite measures (in this case, diabetes) that are theoretically possible without an EHR but infeasible in practice.
- The Billings Clinic case study exemplified the HIT strengths of EHRs to coordinate care and measure its outcomes, in this case for a warfarin/antibiotic alert tied to a warfarin clinic.
- The Kaiser-Portland case study overcame the “free-text dilemma”—that free text, or unstructured information, cannot be readily used for quantitative analyses. They used natural-language processing, in this case for counseling about tobacco use, to capture information in text notes.
- Work at Geisinger focused on reconciling information on the health-problem list (a structured-text field) with information in the visit note (an unstructured-text field).

These case studies, involving five leading provider organizations, highlighted the variety, strengths, and potential challenges of quality indicators associated with EHR data. Although the particulars of each case study were different, the cases presented many common benefits and barriers. Thus, the overall lessons from these providers’ experiences can help guide future efforts to integrate quality measurement into EHR systems.

CASE STUDY #1
Using the EHR to Compile Blood Pressure Measurements

HealthPartners
Minneapolis, Minnesota

James T. Krizak, Shadi Awwad, and Lynne Dancha

Introduction

Hypertension

Controlling high blood pressure (CBP), an important clinical outcome, is now a quality-of-care measure adopted by the National Committee for Quality Assurance (NCQA) for the Health Plan Employer Data and Information Set (HEDIS). NCQA added this measure to its core measurements because hypertension has been a challenge for providers to control—through the mid-1990s, only 27.4 percent of patients with hypertension had their blood pressure under control. Because CBP is included in HEDIS accreditation measures, health plans are focusing more attention on improving this rate.

CBP has also been adopted by Minnesota Community Measurement, a nonprofit entity whose mission is to accelerate the improvement of health by publicly reporting health care information. While NCQA publishes HEDIS measurement results for numerous national health plans, Minnesota Community Measurement publishes quality results for medical groups in Minnesota.

HealthPartners

HealthPartners is the largest consumer-governed nonprofit health care organization in the nation; with 635,000 members, a full-service hospital, more than 50 clinical sites, and nearly 580 practicing physicians.

HealthPartners' own Medical Group, which treats some 40 percent of the health plan's members, fully converted to electronic health records (using the EPIC system) by January 2005. While other medical groups that treat members of HealthPartners have also converted to electronic health records, HealthPartners currently does not have access to their data. Thus our study population represented only 40 percent of the people enrolled in the health plan.

HealthPartners has played major leadership roles in the development both of HEDIS and Minnesota Community Measurement®. While it is now producing numerous quality measures using EHR data—such as for asthma, depression, diabetes, hypertension, children’s health, and adult preventive services—this case study focused on the CBP measure.

Goal

By following HEDIS specifications, a health plan is assured that its results will be accepted by NCQA and also that the results can figure in comparisons with other providers that use HEDIS specifications. Within NCQA’s Quality Compass grade system, there are five bands for any HEDIS measure, corresponding to a health plan’s compliance percentage rate. After achieving certain band grades, a health plan may receive accreditation from NCQA, thereby greatly increasing its credibility in the community. Because blood pressure control is a key indicator among HEDIS measurements, it is a prerequisite to accreditation, and, to gain that status, HealthPartners’ goal is to reach CBP’s highest band.

Sample

HEDIS guidelines require that the population used for CBP include only patients who are 46-85 years old as of December 31 of the measurement year. A subject is considered to have hypertension if there is at least one outpatient encounter, verified by chart review, with an ICD-9 (International Classification of Disease version 9) code of 401 (Essential Hypertension) during the first six months of the measurement year. The patient’s blood pressure is considered to be under control when his or her reading is less than or equal to 140/90.

Measurement

The actual measurement is the percentage of hypertensive patients satisfying the criteria of the above paragraph. Historically, collection of data for the CBP measure required an electronic scan of administrative claims data for an ICD-9 diagnosis code of hypertension, followed by a manual review of the identified patients’ medical records to verify diagnosis and record systolic and diastolic blood pressure readings.

In this case study, HealthPartners followed the HEDIS specifications for the CBP measure—which dictate the minimum size of the random sample—in order to determine the measurement population. The health plan then increased the sample size by 10 percent to allow for diagnostic coding issues on claims data (occasionally, patients without hypertension or with an inappropriate age are identified and need to be removed

from the sample). For 2005 dates of service, HealthPartners' sample for the blood pressure measure was 1,121 patients, identified by looking for a diagnosis of hypertension among claims data.

NCQA requires that the hypertension diagnosis be verified by chart review, so HealthPartners adapted the EHR to do this verification. An electronic scan of structured analytic fields in HealthPartners EHR data found that 464 patients (41%) in the sample population had a substantiated diagnosis of hypertension and a subsequent blood pressure reading less than or equal to 140/90. The remaining 657 patients required a manual review of patient medical records to substantiate the diagnosis of hypertension or to find that blood pressure was under control. These 657 patients either were not HealthPartners Medical Group patients (so they had no HealthPartners' EHR) or they were HealthPartners members whose blood pressure was not under control.

A benefit of using EHR data was elimination of the need for manual review of 464 patient medical records, which saved HealthPartners about \$5,500 (based on a calculated average cost of \$12.00 per record to review). Although such savings per individual are small, when applied over a large population of patients—and especially when combined with other indicators that use EHR data to achieve standard levels—the savings add up.

Verification and Reliability

Using EHR databases in the data warehouse, 464 patients were identified as having hypertension but having their blood pressure under control. To test the data's reliability, we examined the records of a random sample of these patients. Individual records were manually reviewed, using the EHR's front end—the section that auditors use to confirm the diagnosis or find the current blood pressure. Similarly, the date of the most recent blood pressure test and the blood pressure levels recorded from the EHR database were verified. This test was repeated three times because the first and second reviews uncovered errors in the logic that inappropriately contradicted the EHR warehouse. After correcting the logic, a 100 percent match was achieved between records retrieved from the warehouse and the way they appeared in the EHR front end.

Results

HEDIS guidelines have been followed since 2000, with results showing a positive trend in the CBP measure, as shown in the table. The rate represents the percentage of commercial members identified as having both hypertension and their blood pressure under control ($\leq 140/90$).

In HEDIS year 2006, HealthPartners was 0.8 percent away from achieving Band 1—the highest rank in the NCQA ratings, representing the 90th percentile of all HEDIS managed care organizations. The best practice of providers for the CBP measure was CIGNA HealthCare of Colorado, Inc., at 83.7 percent.

Between 2002 and 2003, there was a large jump in the CBP rate, which analysis correlated with HealthPartners' full implementation of its EHR system. During that period the number of EPIC (warehouse that stores EHR data) members increased by 75.8 percent and the number of blood-pressure readings taken from members and stored in EPIC went up 186.7 percent. The table shows these variables' values from services years 2000 to 2006.

Possible inferences from these data:

- Implementation of the EHR results in better documentation, which then leads to an increase in people with hypertension having their blood pressure taken, recorded, and managed, which can then lead to more patients being in control.
- The EHR certainly makes it much easier to find data. Before, auditors had to manually chase down a patient's data, with the possibility of never finding what they were looking for. Now, a query can be done against the EPIC warehouse, and all blood pressures are identified, no matter which HealthPartners clinic was visited. This ready availability of data can lead to a better compliance rate.

Challenges

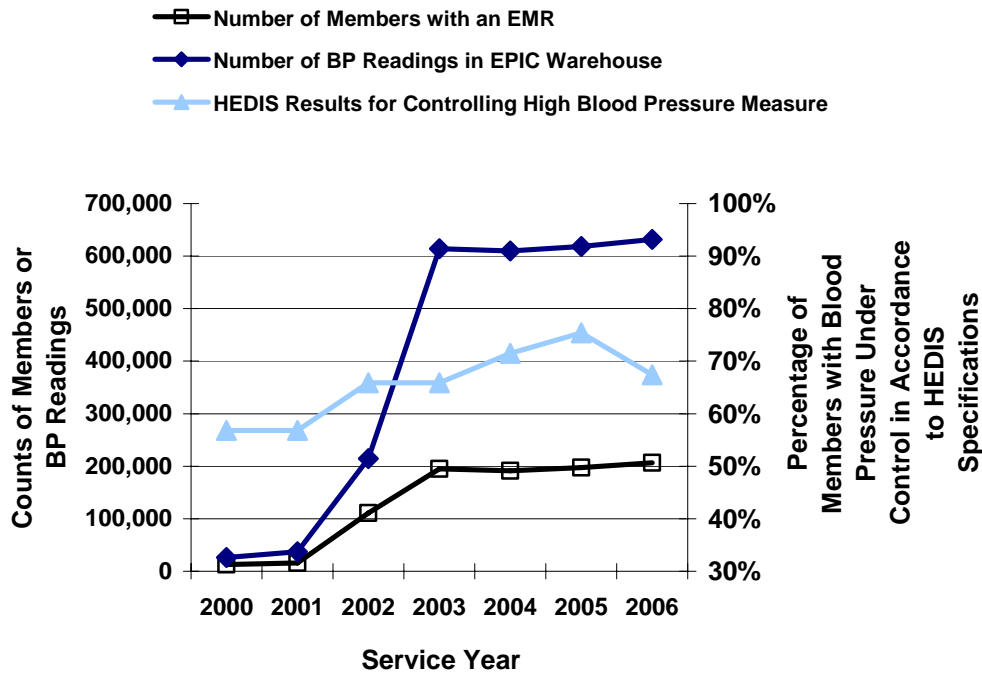
When EHR data were accessed for the first time to report CBP, lack of familiarity with the new warehouse made finding the needed data elements difficult to achieve. EHR extracts were pulled from EPIC and reviewed to verify that the correct codes were being used for the hypertension confirmation and that the blood pressure readings were complete. Still, because review of the EPIC warehouse is an ongoing process, accurate results cannot yet be guaranteed.

Conclusions

The hopes for this movement to EHR are to eliminate the time and cost of pulling data manually and to make the results more robust. When using EHR data, multiple indicators—such as blood pressure, LDL, HDL, HBA1c, BMI, immunization history, and smoking status—can be pulled and reviewed simultaneously. Also, it seems that with the implementation of an EHR system, HEDIS compliance rates may actually increase, as

they appear to do for the Controlling High Blood Pressure measure (see figure below). (The apparent drop in performance in 2006 was, in fact, due to a HEDIS specification change. The requirement for being under control went from $\leq 140/90$ to $< 140/90$, resulting in a surprising impact on this measure's rates.)

Figure 1. HealthPartners EMR Data in Contrast with HEDIS Controlling High Blood Pressure Results



Notes: The 2006 Controlling High Blood Pressure (CBP) rate dropped because of HEDIS specification changes. The requirement for being under control went from $\leq 140/90$ to $< 140/90$.

Although further investigation is needed, the use of EHR data has the potential to increase the quality of care. Not only did HealthPartners' rate for CBP increase through the years, its National Band performance improved as well. This is illustrated in the table below (a lower band is better).

**Table 1. Results for Electronic Health Records (EHRs)
and Blood Pressure (BP) Readings**

HEDIS Measurement Year	Service Year	Number of Members with EHR	Number of BP Readings in EPIC Warehouse	HEDIS Results for Controlling High BP Measure	HEDIS Band
2001	2000	12,493	26,209	56.8%	3
2002	2001	16,030	37,350	56.8%	4
2003	2002	110,940	214,295	65.9%	2
2004	2003	195,031	614,373	65.9%	3
2005	2004	191,158	609,786	71.5%	3
2006	2005	197,220	618,468	75.4%	2
2007	2006	206,548	631,495	67.4%	2

Note: Results for 2006 are the number of EPIC members and BP readings as of 12/18/2006.

Finally, it must be noted that the methods employed in calculating the CBP measure were reviewed and approved by an independent auditor, in accordance with NCQA/HEDIS specifications.

CASE STUDY #2
Optimal Care for Diabetes: Linking Measurement with Improvement

Park Nicollet Health Services
Minneapolis, Minnesota

Nancy Jarvis, M.D., and Shadi Awwad, M.S.

This case study describes the definition, evolution, and use of a composite measure for care of people with diabetes in a clinical setting.

Description of the Study Site

Park Nicollet Health Services—an integrated care system consisting of Park Nicollet Clinic, Park Nicollet Methodist Hospital, Park Nicollet Foundation, and Park Nicollet Institute—in 2003 completed development of a \$60-million health information infrastructure that features a comprehensive electronic health record spanning the hospital and all clinical locations. The vendor for this system is GE Centricity (formerly IDX). The organization's EHR/HIT support includes a single ambulatory/inpatient/home care medical record, a multifaceted clinical decision support system, electronic reminders for patients and clinicians, laboratory/radiology results management, provider/patient communications networks, and computerized physician order entry for the hospital.

Park Nicollet Clinic is one of the largest multispecialty clinics in the United States, providing care in 45 medical specialties and subspecialties. Almost 8,000 employees, including 690 physicians and 270 clinical professionals, are on staff at 25 clinics in Minneapolis and its surrounding suburbs. Park Nicollet Clinic provides care for patients with diabetes through its primary care departments (family medicine and internal medicine), adult endocrine and pediatric endocrine departments, and International Diabetes Center, which provides education for diabetes patients and their families.

Brief Description of the Composite All-or-None Measure

The composite measure is the percent of diabetes patients age 18 through 75 who have achieved hemoglobin A1c <7, LDL cholesterol <100, and blood pressure <130/80; take aspirin daily; and do not use tobacco.

This composite measure is calculated monthly at three levels: (1) individual clinician, (2) site (and department within site, if appropriate), and (3) overall care system.

Rationale for the Measure

In the United States, nearly 21 million people have diabetes. It is a fast-growing epidemic with enormous quality-of-life and public health implications and rapidly increasing health care costs. The above five components of the composite measure were chosen because they are representative of what many organizations, large and small, are using as quality indicators for individuals with diabetes. Attention to this condition is further driven by pay-for-performance and public-reporting initiatives, leading many institutions to choose diabetes as one of the earliest opportunities for management of populations of patients quality improvement, and measurement.

Composite All-or-None Measures

Composite all-or-none measures sum performance across a number of indicators for an individual patient, and credit is given only when all of the indicators have been attained. Interest in such measures has recently been stimulated by the Institute for Healthcare Improvement, the Joint Commission on Accreditation of Healthcare Organizations, and the National Committee on Quality Assurance as an approach that offers a more complete profile of a patient's care for a particular condition (Nolan 2006).

The Minnesota health care community has long been recognized as a leader in innovative practices for delivery and organization of medical care, so it is not surprising that the community's measurements are innovative as well. Thus, care systems in the Twin Cities have been promoting composite all-or-none measures, based on the belief that transformational improvement will happen only when a care system assesses more than individual process and outcome measures—it instead needs to report on patients being up to date on all key indicators. Composite measures appear to accelerate improvement in medical groups with team-based initiatives; it is uncomfortable for clinicians and health care leaders to see personal performance rates of less than 10 percent for composite measures, and reaction is commonly vigorous.

Examples of composite measures in the inpatient setting abound, but composite measures in the ambulatory area are not yet widespread. Composite measures were not feasible before the refinement of the EHR, which now captures key information such as aspirin and tobacco use in structured data fields.

How Composite Measures Drive Workflow Change

Trying to ensure that each patient is up to date for all diabetes care services (hemoglobin A1c, LDL, blood pressure, aspirin, no tobacco use) requires an approach different from that of trying to change a single measure. When measures are bundled into a composite,

all indicators of the measure need to be addressed simultaneously. No solitary tactic (working only on laboratory results, say, or on medications) will work. To move performance on composite measures, team collaboration, planning, workflow redesign, and system reform are necessary.

Evolution of Diabetes Care Measures

One-at-a-time measures of diabetes performance initially came from claims-based data, which provided dates of service for appointments and tests.

Starting in 2000, Park Nicollet's Supporting Best Care Department delivered quarterly reports to primary care managers and individual clinicians. Results were partially blinded, with clinicians seeing only their personal data and comparisons with primary care partners at their own clinic site. At that time, Park Nicollet was still measuring and reporting individual indicators—A1c, LDL, and urine microalbumin values and eye-examination service dates. The phasing-in of the EHR system interrupted the production and distribution of these reports from 2003-2005.

In 2004, vital signs and the medication list were reformatted as structured data in the EHR, and Park Nicollet was able to add blood pressure, aspirin, and tobacco use to the diabetes measure, with the new composite measure being called the "Grand Slam."

Changes in Primary Care

Among Park Nicollet's diabetes patients, primary care clinicians take care of 88 percent, most of whom have type 2 diabetes; physicians in the endocrinology departments take care of the remaining 12 percent. Park Nicollet's major improvement efforts began in the primary care areas by providing measurement feedback and systems support.

The primary care clinicians' diabetes performance measures had been improving over time, but publishing the diabetes data with the composite measures in a partially unblinded fashion across the primary care departments evoked strong reactions from clinicians. In response, primary care administrative leaders implemented better support processes for standard work—tailoring the visit for each patient the day before and, once in the office, getting the patient ready to see the clinician. Now that the basic care processes have been established, Park Nicollet's primary care leaders are directing attention to more aggressive drug therapies for glycemic, lipid, and blood pressure control.

Changes in Specialty Care

After the composite measures were established in the Twin Cities, they were followed by aggressive pay-for-performance incentives, which prompted Park Nicollet administrators

to begin working with the endocrinology departments—whose practice and operations are very different from those of primary care departments. Endocrinologists serve as primary diabetes clinicians for the majority of patients with type 1 diabetes, insulin pumps, end-stage renal disease or transplants, and other significant comorbidities. These specialists, who have complex issues with management of hypoglycemia, for example, argue that it is more difficult for them to achieve the aggressive goals in their patient population. Guidelines for glycemic control, they point out, are established for people with type 2 diabetes, even though the local and national measurement standards reflect the spectrum of diabetes, both type 1 and 2. Park Nicollet clinicians and leaders continue to have lively discussions on how to improve intermediate outcomes and set goals in this group of patients with difficult management issues.

Park Nicollet Performance Over Time

Performance on the composite all-or-none measure over time is shown in the figure below. The graph distinguishes between sites that have instituted an extended laboratory menu (ELM)—the process of obtaining test results from the laboratory immediately before the patient’s visit—and those that have not. This process has created timely face-to-face conversations between the clinician and patient about behavior and drug therapy, allowing the clinician to start or change therapy more successfully in the diabetes population not yet meeting goals.

Composite Measure Specifications

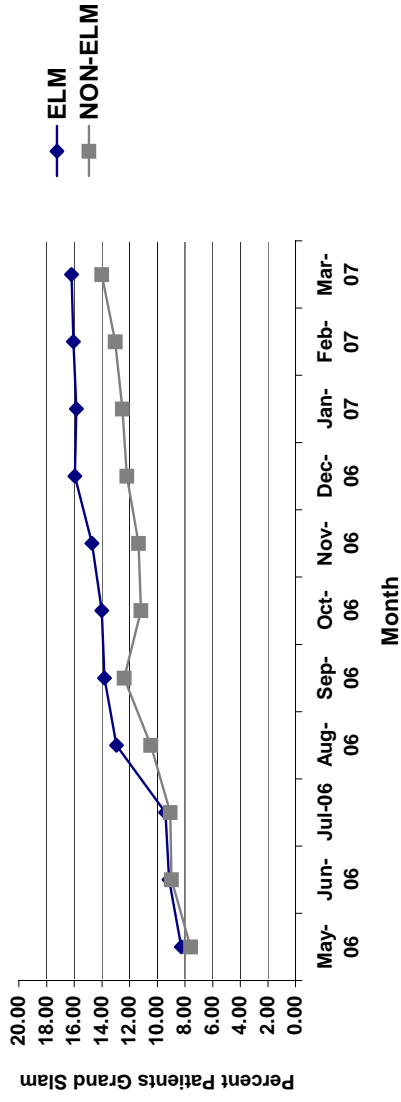
- Denominator: Includes all patients in the diabetes registry who are age 18 through 75 and during the preceding 24 months have been seen at least twice in the ambulatory setting with an ICD-9 diagnosis of diabetes. (At the time the registry was established, Park Nicollet did not have the capacity for automating medication information to use for building the denominator. Also, because the registry was established before the EHR problem list, Park Nicollet could not use the list to enroll patients in the registry.)

- Individual Numerator Measures
 - % hemoglobin A1c < 7 in the past 12 months
 - % LDL cholesterol < 100 in the past 12 months
 - % blood pressure < 130/80 in the past 12 months
 - % daily aspirin use
 - % not using tobacco

Composite Measure: percentage of patients with A1c, LDL, and BP at goals and up to date, who use aspirin daily, and who do not use tobacco (the “Grand Slam”)

	% Patients Managed (Grand Slam)	
	ELM (%)	Non-ELM (%)
May-06	8.25	7.58
Jun-06	9.14	8.98
Jul-06	9.40	9.07
Aug-06	12.95	10.47
Sep-06	13.82	12.40
Oct-06	14.02	11.17
Nov-06	14.73	11.35
Dec-06	15.96	12.20
Jan-07	15.85	12.52
Feb-07	16.06	13.03
Mar-07	16.19	14.02

Figure 2. Extended Laboratory Menu (ELM) vs. NON-ELM Percent Patients Grand Slam



ELM is the process of obtaining test results from the laboratory before the patient's visit. Point of care testing, which also provides test results before the visit, is done in the clinical department by a nurse.

Clinician Assignment

Key to a measure that stimulates quality improvement is the accurate assignment of patients to clinicians. Park Nicollet has used different algorithms over time and currently uses the following rules:

Type 2 Diabetes

- Patients treated by Park Nicollet primary care clinicians are assigned to the clinician directing the majority of the visits for diabetes for that patient.
- Patients who are referred for diabetes consultation to the International Diabetes Center (for education) or to a Park Nicollet endocrinologist remain within the primary care clinician's panel.
- Patients who are treated by non-Park Nicollet clinicians for their diabetes care are assigned to their non-Park Nicollet primary care clinician *and* to the Park Nicollet diabetes clinician (approximately 500 out of 12,000 patients therefore have two assigned clinicians).

Type 1 Diabetes

- Patients who are treated by a Park Nicollet endocrinologist for diabetes care and seen by a Park Nicollet primary care clinician only for non-diabetes visits are assigned to the endocrinologist.
- Type 1 diabetes patients with a non-Park Nicollet primary care clinician and seen by a Park Nicollet endocrinologist will be assigned to the endocrinologist if told to return to him or her within 12 months. If these patients are returned to their non-Park Nicollet clinician with diabetes recommendations, they will be assigned to that clinician.

Data Requirements

- Diabetes measurements are based on standards set by HEDIS. Inclusion in the diabetes cohort for an organization is based on a set of qualifying ICD-9 and CPT Evaluation and Management codes.
- A1c test results are obtained from the data warehouse, having been extracted from the laboratory files within the EHR application (LastWord).
- LDL cholesterol test results are obtained from the data warehouse, having been extracted from the laboratory files within the EHR application (LastWord).

- Systolic and diastolic blood pressures are obtained from the Flowchart Viewer screen of the EHR. (Blood pressure is taken only from clinic readings, not hospital readings.)
- Tobacco use status is obtained from the Flowchart Viewer screen of the EHR. (Credit is only given if the patient is a nonsmoker.)
- Daily aspirin use (any amount) is obtained from the Outpatient Medications screen in the EHR.

Adding Laboratory Values that Originate Outside Park Nicollet

Because Park Nicollet has many patients who have testing done at outside laboratories or are participating in research studies that use reference laboratories, a critical validity issue for measurement is the documentation of studies done outside the Park Nicollet setting. This external testing process pertains especially to patients seen in specialty departments. To deal with the potentially missing data, Park Nicollet has added a structured flowsheet that captures external sources of laboratory data and can be queried for test results and dates of service.

Conclusion

Many opportunities and challenges developed during the design, building, and implementation phases of this composite all-or-none measure. The issues that Park Nicollet continues to address for improving performance on the measure include standard work for: planning between visits, tailoring the visit for each patient the day before, getting the patient ready to see the clinician in the office, organizing the components of the clinician visit, assuring that needed information is at hand for prescription changes, and having processes in place for follow-up on testing and other outcomes of the visit.

Composite measurement has dramatically altered the approach to quality improvement at Park Nicollet. The organization understands that infrastructure and teamwork are needed to raise the bar on performance.

CASE STUDY #3
Warfarin/Antibiotic Rule for the EHR

Billings Clinic
Billings, Montana

Patricia Coon, M.D.

Goal

The overall goal of this case study is to improve patient safety by reducing adverse drug responses—in particular, hemorrhagic events caused by warfarin/antibiotic interactions.

Rationale

The Warfarin/Antibiotic Rule, an innovative care indicator that takes advantage of the unique capabilities of the EHR and related HIT, is designed to improve communication between the health care provider who is ordering an antibiotic for a patient on warfarin and the pharmacist who is managing the patient’s chronic anticoagulation therapy. A number of commonly prescribed antibiotic, antiviral, and antifungal medications have significant interactions with warfarin, thereby increasing the risk for bleeding. When these medications are ordered, more frequent monitoring of a patient’s prothrombin time (INR)—a standardized way of monitoring blood-clotting—is required to avoid adverse events. EHR prompting may also reduce risk by altering the antibiotic-prescribing habits of health care providers. Choosing an antibiotic that produces little or no interaction with warfarin, rather than one with high interaction, minimizes the risk of complications and the need for more frequent monitoring. Our organization believes that the rulemaking capacity of the EHR provides a viable alternative to existing electronic alerts on drug interactions, which frequently cause “alert fatigue.”

Description

When a provider in our health care system attempts to electronically prescribe a high-risk antibiotic and the patient has warfarin listed on his or her medication profile, one of two text boxes pops up on the screen:

Text Box Pop-Up for Patients Followed by the Anticoagulation Clinic

This text box alerts the provider to a potentially moderate- to high-risk drug/drug interaction. It also states that if the clinician persists in prescribing this antibiotic, an alert will be automatically sent to the pharmacist in the anticoagulation clinic. At this point,

the provider can launch a screen that lists warfarin/antibiotic interactions, ranked in order from no risk to high risk, so that he or she may chose a different antibiotic that is equally effective but causes little or no warfarin interaction. If the provider indeed prescribes such an antibiotic, no alert is “fired.” In the case when such an alert is fired and received, the pharmacist electronically documents the action taken, using a drop-down box format.

Requirements for the rule to fire: the patient has warfarin on his or her medication profile, the antibiotic ordered has moderate to high interaction with warfarin, and the patient is followed by the anticoagulation clinic.

Text Box Pop-Up for Patients Not Followed by the Anticoagulation Clinic

This text box alerts the provider to a potentially moderate- to high-risk drug/drug interaction and encourages him or her to get a prothrombin time/INR measurement of the patient in three days if the medication is prescribed. This alert also allows providers to launch a screen that lists warfarin/antibiotic interactions, ranked in order from no risk to high risk, thereby enabling them to choose a medication with minimal interaction. If an antibiotic with moderate to high risk is prescribed, providers have the option to print out a patient information sheet that instructs them to call the primary care provider for further management. This text box pop-up was developed for use by our rural health care providers and urgent care providers, who see patients not followed by our anticoagulation clinic.

Requirements for the rule to fire: the patient has warfarin on his or her medication profile, and the antibiotic ordered has moderate to high interaction with warfarin.

Development and Refinement

To develop the Warfarin/Antibiotic Rule, an ad hoc committee comprising individuals from Billings Clinic’s Information Services (IS), Electronic Medical Record Committee, and Pharmacy Department was formed. Committee members included a pharmacist from the Anticoagulation Clinic, two primary care providers, a primary care office coordinator, and the key personnel from IS. Committee members formulated and agreed on the parameters for the rule, and IS staff wrote the rule and implemented it into the EHR. The rule was then tested for reliability in a nonclinical environment.

In September 2006, the Warfarin/Antibiotic Rule was partially activated in the clinical environment. At that time, pharmacists in the Anticoagulation Clinic were alerted when any patient on warfarin therapy was placed on a nontopical antibiotic, antiviral, or antifungal agent, regardless of the degree of interaction. The clinic received up to 20 such alerts daily, which required a significant time commitment and interfered with workflow.

Several primary care providers agreed to pilot the new rule, which had been programmed to fire pop-up alerts if they ordered an antibiotic with anywhere from low to severe warfarin interaction.

Two months after implementing the Warfarin/Antibiotic Rule, the ad hoc committee met with representative individuals piloting the alert. There was consensus on three issues:

- 1) The rule should only fire when an antibiotic, antiviral, or antifungal with a moderate to severe warfarin/drug interaction is ordered. No clinical action is needed for medications with little or no interaction. This modification would significantly reduce the workload for the Anticoagulation Clinic staff.
- 2) The Pharmacy Department would prepare a table that rated warfarin and antibiotic/antiviral/antifungal interactions as either none, low, moderate, or severe. Health care providers who had begun prescribing a high-risk drug could launch this table directly from the pop-up box alert in order to search for an alternative—i.e., a medication that is equally effective but has little or no interaction with warfarin.
- 3) Two pop-up alerts were needed—one for patients followed by the Anticoagulation Clinic and another for those whose anticoagulation is managed by primary care providers.

Primary Outcome Measures

The effect of the Warfarin/Antibiotic Rule on the following outcome measures will be tracked quarterly:

- Percent of elevated INR excursions—i.e., $\text{INR} > 6$ —caused by warfarin-antibiotic interaction.
- Number of antibiotic-induced elevated INR excursions per patient followed by the Anticoagulation Clinic.

Data Sources

EHR electronic medication profile; EHR inbox; and Anticoagulation Clinic electronic monitoring form—an e-form for all patients followed by the clinic with the patient's name and anticoagulation monitoring information.

Reach

The rule affects all health care providers in the system who prescribe antibiotics electronically in the outpatient setting, which includes the outpatient clinic, urgent care clinics, emergency room, and outpatient surgery. The rule will not fire in the inpatient setting because pharmacists already check all medications ordered there for drug/drug interactions.

Challenges

- Loss of health care provider autonomy—i.e., the alert is automatically fired to the Anticoagulation Clinic once the antibiotic has been e-prescribed. Although this was a concern brought up by the ad hoc committee, it has not proven to be an issue.
- Restructuring of workflow process for pharmacists in Anticoagulation Clinic.

Current Status

Development of the electronic Warfarin/Antibiotic Rule is now complete. The rule was piloted in several primary care provider offices from September 2006 to March 2007. Systemwide implementation—that is, the training of all physicians and office nurses with access to our EHR—was finished by March 2007. Independent of providers, pharmacists in the Anticoagulation Clinic have been receiving alerts on clinic patients since September 2006.

Principal Findings

Since implementation, the alert has been fired an average of 3–4 times a day for a panel size of approximately 1,300 patients. The percent of patients with elevated INR excursions ($\text{INR} > 6$) caused by warfarin-antibiotic interaction has decreased from 0.20 to 0.13. Preliminary reports show that the number of antibiotic-induced elevated INR excursions per patient followed by the Anticoagulation Clinic has been reduced by about 25 percent.

Clinician Comments

The pharmacists in the Anticoagulation Clinic believe that this rule will improve patient safety throughout the organization. Because the clinic receives about 2–3 alerts per day, they do not think the rule appreciably increases daily pharmacy workload or interferes with workflow. Pharmacists suggest that it may actually improve workflow by decreasing telephone calls from health care providers and patients who are alerting them about the start of antibiotic therapy. Pharmacists note that patients are surprised and pleased to receive a telephone call from the Anticoagulation Clinic within hours of obtaining the antibiotic prescription.

Primary care providers (PCPs) and office nurses who piloted the rule do not believe that it increases office workload or interferes with workflow. The providers feel that the pop-up boxes are clear and easy to use; and, because the rule fires so infrequently, “alert fatigue” from the boxes is not a concern. Providers enthusiastically support use of the rule to automatically alert the Anticoagulation Clinic when a moderate- to high-risk antibiotic is ordered; loss of PCP autonomy is not perceived to be an issue. Upon seeing the pop-up box alert, some PCPs have used the warfarin/antibiotic interaction list to choose a different antibiotic that is equally effective but has little or no warfarin interaction. The PCPs and office nurses believe that the rule will enhance patient safety.

Conclusions

Implementation of an EHR-based Warfarin/Antibiotic Rule has a positive effect on provider-to-provider communication and may reduce the risk of adverse outcomes—e.g., hemorrhagic events caused by warfarin/antibiotic interactions. This rule relies on traditional components of an EHR—the electronic medication profile, for example. Therefore, it is plausible that other health care organizations with comprehensive EHRs might see these systems’ rulemaking capacities as capable of improving communication among health care providers and enhancing patient safety.

CASE STUDY #4
Assessing Clinician Adherence to Smoking-Cessation Guidelines
Using MediClass: An Innovative Natural-Language Processing Tool

Kaiser Permanente of the Northwest
Portland, Oregon

Dean F. Sittig, Ph.D., and Brian L. Hazlehurst, Ph.D.

Introduction

Many quality measures, such as HEDIS, use smoking-cessation counseling as a key indicator of quality of care (Solberg 2003, Chin 1997). There are three main reasons: smoking's status as the leading preventable cause of death in the United States (CDC 2003, HHS 2004, Mokdad 2004); smoking cessation's aid in reducing the risk of premature death (Wewers 2003); and the fact that direct counseling by the patient's primary care provider is one of the most successful methods of getting patients to stop smoking, with a quit rate of up to 10 percent (Prochazka 2000, Cokkinides 2005). Current practice is to use patient surveys (by mail or telephone) to measure clinician performance, though the results obtained from the two survey methods can be quite different (Solberg 2003). We believe that medical-record review is a better alternative; it is more accurate and reliable and less bothersome to patients. Also, in a large integrated group-practice setting, patients are often seen by more than one health care provider. Consequently, it is critical that all providers accurately record their findings, advice, actions, and plans in the medical record so that the next providers can pick up where they left off. Therefore we were interested in developing an EHR-based measure, based solely on the data contained in our state-of-the-art electronic medical record system, of how well our clinicians were adhering to the smoking-cessation counseling guidelines described below (Chin 1997).

Measuring the 5As of Smoking-Cessation Care

The nationally recommended treatment model (Dolin 2006) involves five steps, or "the 5As": Ask, Advise, Assess, Assist, and Arrange follow-up. Although some of these steps are easily coded into the EHR (e.g., identification of smoking status or prescriptions for smoking-cessation drugs), others are typically recorded free-text style in the progress notes or the patient instruction portions of the medical record.

Below is a table of the 5As, their operational definitions, and an example of free-text notation in the EHR.

Table 2. The “5As” Recommended by the Current U.S. Public Health Service Clinical Practice Guideline for Tobacco Treatment and Prevention

5A Step	Operational Definition	Example in Free-Text Section of EHR
Ask	Identify tobacco-user status at every visit.	“Patient smokes 1ppd.”
Advise	Advise all tobacco users to quit.	“It is important for you to quit smoking now.”
Assess	Determine the patient’s willingness to make a quit attempt.	“Patient not interested in quitting smoking.”
Assist	Aid the patient in quitting.	“Started patient on zyban.”
Arrange	Schedule a follow-up contact, in person or by telephone.	“F/u in 2 wks for quit progress.”

Design and Development of MediClass

The MediClass system was built from open-source and freely available software components. It uses three distinct information and communication system technologies:

- The Health Level 7 (HL-7) Clinical Document Architecture (CDA) (Dolin 2006)
- Natural-language processing (NLP) techniques
- Knowledge-based systems.

MediClass currently employs a customized version of the CDA that simplifies classification processing while permitting compatibility with this emerging standard for health record data.

MediClass combines NLP techniques with knowledge-based systems technology. Using NLP knowledge modules, MediClass maps the contents of each encounter to a controlled set of clinical concepts based on (1) phrases detected in free-text sections and (2) codes detected in structured sections of the medical record. This knowledge was encoded into the MediClass system as an application-specific “knowledge module.”

Testing of NLP Knowledge Module

The project requested the electronic health records of about 1,000 known smokers at each of four institutions. These EHRs included relevant data from single office visits with primary care clinicians.

In preliminary work, 500 records—125 randomly selected from each of the four sites—were coded both by trained chart abstractors and MediClass (MC). After disagreements (such as misclassifications) between MC and the human raters were assessed, MediClass was improved and rerun to validate the revised system. The records used in this preliminary work were then removed from the data pool, leaving 875 records from each health plan.

Validation of the MediClass System

Our preliminary work (described above) as well as previous studies showed that several of the “As” are infrequently found in the data. Therefore, the validation study used a sample of records composed both of random and “enriched” portions (the latter located by the MediClass system). Because the final size of the enriched portion of our sample was 77 records, the rest of our sample (423 records) was then randomly drawn from the remaining data pool, stratified by health plan. Thus the preliminary work’s final sample of 500 records contained 125 records from each of the four health plans.

Four trained medical record reviewers coded all 500 records. These abstractors were trained to look for evidence of documentation of the 5As of smoking-cessation counseling, and they were asked to identify whether each “A” was absent or present in each record. The sites anonymously submitted the results of this work to our coordinating center.

Results of the Validation

To analyze these data we created a “gold standard” based on the majority opinion of the human raters (i.e., when three or four of the raters agreed on the presence or absence of a particular A). We then computed the accuracy of MediClass against this standard.

MediClass agreed with the gold standard 91 percent of the time. Point estimates of sensitivity for each of the first four As (the fifth A could not be compared, because of the low frequency of its occurrence) were found to be 0.97, 0.68, 0.64, and 1.0, respectively. Point estimates of specificity were 0.95, 1.0, 0.96, and 0.82.

Impact of MediClass Use on Physicians

A clinical trial involving four disparate health plans used MediClass to generate quarterly feedback reports to individual doctors. These reports were based on information contained both in the free-text and coded sections of electronic health records in routine clinical use. A preliminary analysis found improved adherence on some of the key measures of smoking-cessation care. A complete report on the trial’s findings is in preparation.

Lessons Learned from MediClass

Over the course of the clinical trial we learned many important lessons regarding the use of NLP tools to analyze both the coded and free-text data contained in EHRs. These lessons included the following:

- The practice of medicine is constantly changing, as are the EHRs that clinicians use to document their actions. Thus, while many people think that once an EHR is installed you are “home free,” it turns out that changes in the EHR infrastructure, or the entire EHR, are not uncommon (Campbell 2006). In fact, two of the health plans replaced their EHRs during the four-year study period.
- The NLP solution was more consistent than the human chart abstractors, although this solution has both pluses and minuses. For example, the NLP solution never gets tired reading the notes, and never varies in its interpretations of the data. Human coders on the other hand continue to make new and different errors over time, and can exhibit significant variability among coders.
- As clinicians become more acclimated and knowledgeable about using EHRs, they may begin recording less and less information in the free-text portions of the record. We expect they may record less because they feel that the coded data from the clinical laboratory and their orders adequately document their findings and actions. Unfortunately, secondary reviewers of the chart, be they other clinicians, or medical chart abstractors, can not always determine the original clinician's interpretation of the data or their motivation behind the resulting orders.
- The use of “automatic,” or “canned,” phrases can wreak havoc on an NLP system and challenge our ability to interpret the text, whether by machine or by humans. Specifically, this machine-generated text is "not natural" and is often formatted and expressed differently and in much greater volume than routine, human-generated text. On the other hand, if the same "canned text" phrases are used repeatedly, the NLP solution can be trained to understand these phrases very easily.
- Clinicians use EHRs in an ad hoc manner. In other words, different clinicians use diverse system features in many different ways. This variation leads to the information being recorded in many different ways and in many different physical locations within the record. The fact that both humans and NLP solutions need to look for this information in these disparate locations makes it difficult to interpret the meaning of information entered into the system.

Future Enhancements and Plans for MediClass

Since we developed the knowledge modules required to identify smoking-cessation counseling behaviors, we have created several more knowledge modules capable of:

- Finding evidence of adverse reactions to vaccinations (Hazlehurst 2005);
- Identifying clinical classifications of diabetic retinopathy that are not supported by regular ICD-9 coding schemes (Smith 2005); and
- Identifying patients with a family history of cancer (Hazlehurst).

Conclusion

Informative and essential data within the free-text portion of clinicians' notes and other text portions of the EHR are unavailable to most methods of automated health care assessment. Because of the clinical value of the EHR narrative, and the poor acceptance thus far of structured data entry, one might therefore think it unlikely that wholesale replacement of the narrative with structured data entry will succeed. But MediClass demonstrates the feasibility of an automated coding system for processing the *entire* EHR, enabling assessment of smoking-cessation care delivery. Such a system can be similar in accuracy to that of trained human coders. Systems like MediClass can help bridge the gap between the promise and the realization of EHRs' value.

CASE STUDY #5
Problem List Diagnosis Reconciliation

Geisinger Health System
Danville, Pennsylvania

Mark Selna, M.D.

The Problem to Overcome, the Opportunity to Realize

If information on diagnoses, conditions, or risk factors does not exist in the patient’s EHR, does not exist in any other structured and unambiguous context, or is inconsistent within the record itself, its utility with regard to effective decision support is limited—that is, the information is not accurate, current, or readily available.

The diagnoses with the greatest validity continue to be those that the treating provider has actively added to the patient’s “Problem List.” But, because the documentation workflow inherent in most EHRs does not require reconciliation of encounter diagnoses with those on the patient’s Problem List, and because EHR Problem Lists are usually not as visually evident as those within paper medical records, most EHR data sets are prone to having incomplete Problem Lists as well as significant discrepancies between the patients’ encounter diagnoses and their Problem List diagnoses.

For resource-intensive or high-morbidity clinical care processes, such as disease-specific case management programs, the targeting specificity that can be achieved with a high-integrity Problem List is invaluable. Thus, to assess the integrity of the Problem List diagnoses—both with regard to accuracy as well as internal consistency (i.e., relative to other representations, such as encounter diagnoses, in the record)—quantitative performance measures are required.

Indicator Description

The percentage of patients with selected “high-impact” chronic-disease diagnoses, based on ambulatory encounters with providers, that have such diagnoses on their “Problem List.”

Indicator Specifications

The eligible population comprises patients who must:

- 1) Have had at least two completed ambulatory encounters with a Geisinger Health System (GHS) Primary Care Provider (PCP) within the prior 24 months
- 2) Have at least one of the high-impact chronic diseases associated with their completed ambulatory PCP encounters or active on their Problem List.

Denominator: The combined count of all unique (patient-level) high-impact ambulatory encounter diagnoses from the prior 24 months.

Numerator: The combined count of all unique (patient-level) high-impact ambulatory encounter diagnoses that have the corresponding high-impact diagnosis active on the patient's Problem List.

Required Data (for each eligible patient)

- Current PCP designation (Geisinger or non-Geisinger)
- Ambulatory encounter diagnoses from the prior 24 months (PCP and non-PCP encounters)
- Current Problem List diagnoses.

Required Data Sources (EpicCare)

- Table: Patient (Field: Current PCP Provider ID)
- Table: Patient Encounter Diagnosis (Field: ICD-9 Code)
- Table: Problem List (Fields: ICD-9, Status)

Supplemental Information

- Table: Provider Master File (Fields: Provider ID, Employment Status)
- Table: High-Impact diagnoses (descriptions, ICD codes)

Indicator Development

The primary challenges in developing this indicator were the many arbitrary decisions required in defining the specifications. For example, high-impact chronic-disease diagnoses were primarily chosen in accordance with their population-wide prevalence or medical expenses, as opposed to their patient-specific burdens.

Another common challenge was deciding which disease-specific ICD code sets would yield the most accurately diagnosed target population—the methodological bias was

to favor third-party specifications (e.g., CMS Physician Group Practice Demonstration; applicable HEDIS specifications). Lastly, in order to better ensure that the population profiled by the indicator was one that was receiving ongoing management by the Geisinger Health System (i.e., at least one follow-up opportunity for the PCP to “reconcile” the Problem List), only patients who had had at least two completed ambulatory encounters with a Geisinger PCP within the prior 24 months were included. (Patients who expired during the reporting period were not excluded.)

The high-impact chronic-disease diagnoses chosen were congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), diabetes, depression, coronary artery disease (CAD), and hypertension (HTN).

Indicator Evaluation

Using the specifications described above, independent diagnosis-specific populations (denominators) were identified. Each population was then stratified according to two variables: (1) the presence or absence of that diagnosis on the patient’s Problem List; and (2) the number of high-impact disease-specific PCP ambulatory encounters. (Note that many patients had multiple high-impact chronic-disease diagnoses and, as such, were included in several independent denominator populations.)

As shown in the attached table, the indicator revealed that the proportion of target-population patients with the required ICD code on the Problem List varied from a high of 89 percent to a low of 59 percent, depending on the high-impact chronic-disease diagnosis. However, it is not clear from this case study how much the ambulatory visit frequency threshold or the length of the performance period affected the utility of the indicator.

Indicator Feasibility and Generalizability

Because this indicator relies only on traditionally codified data (i.e., ICD codes), it is feasible to administer. The primary caveat is with regard to a provider’s encounter-based Problem List coding practices. Specifically, unless the Problem List diagnoses, as with the encounter diagnoses used to generate a billing claim, are represented by (or linked to) specific structured diagnosis data, reliable reproducible correlations are not possible.

For this indicator to be generalizable, it would need to support a provider’s clinical-management or business objectives. For example, given the recent expansion of diagnosis-based quality reporting prompted by “pay-for-performance” initiatives, some

providers might rely more heavily on their Problem Lists to develop disease-specific registries and tracking programs.

The operational questions raised by this indicator involved the potential use of less manual means of populating the Problem Lists. Specifically, one could ask providers to consider whether any of the following three options were clinically appropriate and acceptable to them: (1) auto-populate the Problem List with every high-impact chronic-disease ambulatory encounter diagnosis; (2) impose a “hard-stop” alert that presents at the time of encounter-diagnosis coding to notify the provider of an opt-out “auto-population” of the Problem List; or (3) require providers to retrospectively review and reconcile lists of their patients whose high-impact chronic-disease encounter diagnoses do not appear on their Problem Lists.

This measure has served as the impetus to engage in additional analyses and “clean-up” work (such as eliminating duplicate diagnoses, eliminating or explicitly mapping standard diagnosis codes, or resolving mutually exclusive diagnoses).

Table 3. Diagnosis Specificity for Selected Chronic Diseases—Problem List vs. Encounters
 Reporting Period: 09/01/2003 to 08/30/2005

Problem List Diagnosis	Total Patients		Frequency Distribution of Ambulatory Encounters									
	#	%	0	1	2	3	4	5	6	>6	Total	
CHF	Y	74%	36%	11%	10%	8%	7%	6%	5%	17%	100%	
	N	26%		56%	19%	10%	6%	4%	2%	3%	100%	
COPD	Y	61%	41%	15%	10%	8%	7%	4%	4%	11%	100%	
	N	39%		64%	19%	8%	4%	2%	1%	2%	100%	
Diabetes	Y	59%	24%	8%	8%	7%	8%	8%	7%	30%	100%	
	N	41%		17%	11%	9%	9%	9%	9%	36%	100%	
Depression	Y	70%	41%	16%	13%	9%	7%	4%	3%	7%	100%	
	N	30%		55%	24%	11%	5%	2%	1%	2%	100%	
CAD	Y	78%	20%	16%	12%	11%	11%	8%	6%	16%	100%	
	N	22%		57%	20%	9%	5%	3%	2%	4%	100%	
HTN	Y	89%	18%	7%	5%	4%	3%	3%	3%	57%	100%	
	N	11%		57%	22%	10%	6%	3%	1%	1%	100%	

FIVE CASE STUDIES: CONCLUSIONS

In Park Nicollet’s composite measure that spanned the range of diabetes care, in Billings Clinic’s warfarin/antibiotic alert, and in the other case studies, it was striking how performance measures became much more clinically relevant when they were HIT-based.

But major barriers in these e-indicators’ conceptualization and development, such as the *validity* of data extracted from the EHR, became apparent as well. Many of the case studies noted the challenge of correctly defining a specific patient population—a critical prerequisite to accurate assessment—using ICD-9 classification codes; and two studies (Park Nicollet and HealthPartners) highlighted issues involving incorporation of data from patient-related actions (e.g., procedures, visits, or referrals) that occurred outside the EHR system.

Most case studies implemented workable solutions to these validity concerns. For example, providers used manual chart-review comparisons, more specific diagnostic codes (such as SNOMED), and specialized structured flow sheets to integrate outside data. These methods helped to promote standardized and comprehensive patient coding, which is critical to any EHR-derived quality measure.

Another major barrier was the *reliability* of the EHR-extracted data, which may not be collected and recorded consistently across the patient population. Many case studies pointed out the difficulty of coding data about patients, diagnoses, and procedures in any uniform way; it can be challenging just to record information consistently within one patient’s EHR.

Two of the case studies addressed concerns with standardization through opposite data-collection strategies. Geisinger examined the reliability of using a patient’s Problem List to automatically identify “high-impact” cardiovascular patients. This case study suggested the possibility of implementing algorithms to ensure that patients are not incorrectly excluded from quality measures. Kaiser, on the other hand, examined the potential of using free-text language-recognition software to translate physicians’ notes into the equivalent of structured data fields.

Prior to their implementation, providers expressed concern that the measures would hinder *workflow* or suffer from staff resistance. In actuality, few such events

occurred. On the contrary, the case studies indicated that EHR systems improved workflow by automating key communications between staff and by improving patient-record accessibility across different clinics. Even where staff initially resisted the new systems, additional staff-support infrastructure and departmental leadership dramatically increased both measure compliance and overall performance.

Measures that translated established quality indicators had the easiest transition into EHR. These measures usually had clearly defined specifications, after all, and in many cases the corresponding data were already being compiled through manual chart reviews. Consequently, the measures readily lent themselves to comparisons against past performance, particularly on dimensions such as cost savings, compliance, and provider quality.

Measures incorporating or evaluating HIT-specific features, such as automated alerts and free-text analysis, tended to be more specialized and not as easily incorporated into other systems. Nonetheless, most of the providers were confident that the concepts could be adapted to different EHR system types. Additionally, specific actions were taken to standardize specifications as much as possible, made easier by the fact that populations were increasingly being defined through standard coding methods. Thus many patient-specific data elements necessary to calculate an indicator were likely to be accessible in other systems—though a high degree of technical skill would be needed to implement, maintain, and evaluate indicators incorporating HIT-specific features.

All of the problems that the case-study measures assessed are amenable to performance improvement, and this improvement is made possible by the EHR. Even in the Geisinger example, one can “force” concordance between the Problem List and provider-encounter diagnoses through changes to the EHR.

The success of the case studies’ providers in implementing EHR-based quality measures demonstrates that such measures are worth pursuing, despite the challenges of ensuring the validity and reliability of data, efficient workflow, and staff support. The cases illuminate the paths that need to be followed to move effectively into 21st-century quality measurement. Although using established indicators can facilitate such a transition, deploying and measuring HIT-specific features offers valuable benefits that highlight EHRs’ unique contributions to quality.

REFERENCES

- Ash, J. S., M. Berg, and E. Coiera (2004). "Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-Related Errors." *Journal of the American Medical Informatics Association* 11(2):104–12.
- Bates, D. W., M. Cohen, L. L. Leape, J. M. Overhage, M. M. Shabot, and T. Sheridan (2001). "Reducing the Frequency of Errors in Medicine Using Information Technology." *Journal of the American Medical Informatics Association* 8(4):299–308.
- Berger, R. G., and J. P. Kichak (2004). "Computerized Physician Order Entry: Helpful or Harmful?" *Journal of the American Medical Informatics Association* 11(2):100–03.
- Campbell, E. M., D. F. Sittig, J. S. Ash, K. P. Guappone, and R. H. Dykstra (2006). "Types of Unintended Consequences Related to Computerized Provider Order Entry." *Journal of the American Medical Informatics Association* 13(5): 547–56.
- CDC (Centers for Disease Control and Prevention) (2003). "Cigarette Smoking-Attributable Morbidity—United States, 2000." *Morbidity and Mortality Weekly Report* 52:842–44.
- Chin, H. L., and M. Krall (1997). "Implementation of a Comprehensive Computer-Based Patient Record System in Kaiser Permanente's Northwest Region." *MD Computing* 14(1):41–45.
- Cokkinides, V. E., E. Ward, A. Jemal, and M. J. Thun (2005). "Underuse of Smoking-Cessation Treatments: Results from the National Health Interview Survey, 2000." *American Journal of Preventive Medicine* 28(1):119–22.
- Dolin, R. H., L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, and A. Shabo Shvo (2006). "HL7 Clinical Document Architecture, Release 2." *Journal of the American Medical Informatics Association* 13(1):30–39.
- Hazlehurst, B., J. Mulooly, A. Naleway, and B. Crane (2005). "Detecting Possible Vaccination Reactions in Clinical Notes." *AMIA Annual Symposium Proceedings*, 306–10.
- Hazlehurst, B., E. Harris, and W. Hu (In press). *Extracting Family Cancer History from Text Notes in the Electronic Medical Record*. HMO Research Network Annual Conference (March, 2007) Portland, Ore.
- HHS (U.S. Department of Health and Human Services) (2004). *The Health Consequences of Smoking: A Report of the Surgeon General*. Atlanta, Ga.: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- Koppel, R., J. P. Metlay, A. Cohen et al. (2005). "Role of Computerized Physician Order Entry Systems in Facilitating Medical Errors." *Journal of the American Medical Association* 293(10):1197–1203.
- Mokdad, A.H., J. S. Marks, D. F. Stroup, and J. L. Gerberding (2004). "Actual Causes of Deaths in the United States, 2000." *Journal of the American Medical Association* 291(10):1238–45.
- Nolan, T., and D. M. Berwick (2006). "All-or-None Measurement Raises the Bar on Performance." *Journal of the American Medical Association* 295(10):1168–70.
- Prochazka, A. V. (2000). "Developments in Smoking Cessation." *Chest* 117(4 Suppl 1): 169S–175S.
- Smith, D., E. S. Johnson, A. Russell, B. Hazlehurst, C. Muraki, G. Nichols, H. Kan, and J. Brown (2005). "Associations Among Visual Impairment, Stage of Diabetic Retinopathy, and Quality of Life (EQ-5D)" (abstract). *Value Health* 8(3):365.

- Solberg, L. I., J. A. Hollis, V. J. Stevens, N. A. Rigotti, V. P. Quinn, and M. Aickin (2003). "Does Methodology Affect the Ability to Monitor Tobacco Control Activities? Implications for HEDIS and Other Performance Measures." *Preventive Medicine* 37(1):33–40.
- Tang, P. C., M. Ralston, M. Arrigotti, L. Qureshi, and J. Graham (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(1):10–15.
- Weiner, J. P., T. Kfuri, K. Chan, and J. Fowles (2007). "e-Iatrogenesis': The Most Critical Unintended Consequence of Computerized Physician Order Entry (CPOE) and Other Health Information Technology." *Journal of the American Medical Informatics Association* 14(3):387–88.
- Wewers, M. E., F. A. Stillman, A. M. Hartman, and D. R. Shopland (2003). "Distribution of Daily Smokers by Stage of Change: Current Population Survey Results." *Preventive Medicine* 36(6):710–20.

RELATED PUBLICATIONS

Publications listed below can be found on The Commonwealth Fund's Web site at www.commonwealthfund.org.

[*Quality Matters*](#). Newsletter.

[*Physician Organization in Relation to Quality and Efficiency of Care: A Synthesis of Recent Literature*](#) (April 2008). Laura Tollen.

[*The Feasibility and Value of New Measures Showing Patterns of Quality for Patients with Three Chronic Conditions*](#) (January–March 2008). Stephen M. Davidson, Michael Shwartz, and Randall S. Stafford. *Journal of Ambulatory Care Management*, vol. 31, no. 1.

[*Current Approaches to Improving the Value of Care: A Physician's Perspective*](#) (December 2007). Howard B. Beckman, Thomas Mahoney, and Robert A. Greene.

[*A High Performance Health System for the United States: An Ambitious Agenda for the Next President*](#) (November 2007). The Commonwealth Fund Commission on a High Performance Health System.

[*Patient-Centered Care: What Does It Take?*](#) (October 2007). Dale Shaller.

[*Creating Payment Systems to Accelerate Value-Driven Health Care: Issues and Options for Policy Reform*](#) (September 2007). Harold D. Miller.

[*Value-Driven Health Care Purchasing: Four States that Are Ahead of the Curve*](#) (August 2007). Sharon Silow-Carroll and Tanya Alteras.

[*An Analysis of Leading Congressional Health Care Bills, 2005–2007: Part II, Quality and Efficiency*](#) (July 2007). Karen Davis, Sara R. Collins, and Jennifer L. Kriss.

[*Health Care Opinion Leaders' Views on the Quality and Safety of Health Care in the United States*](#) (July 2007). Katherine K. Shea, Anthony Shih, and Karen Davis.

[*Measuring Hospital Performance: The Importance of Process Measures*](#) (July 2007). Anthony Shih and Stephen C. Schoenbaum.

[*Aiming Higher: Results from a State Scorecard on Health System Performance*](#) (June 2007). Joel C. Cantor, Cathy Schoen, Dina Belloff, Sabrina K. H. How, and Douglas McCarthy.

[*Evidence-Informed Case Rates: A New Health Care Payment Model*](#) (April 2007). François de Brantes and Joseph A. Camillus.

[*Why Not the Best? Results from a National Scorecard on U.S. Health System Performance*](#) (September 2006). The Commonwealth Fund Commission on a High Performance Health System.