FINAL REPORT

Using Precision Performance Measurement to Conduct Focused Quality Improvement

Principal Investigator: David W. Baker, MD, MPH

Team Members: Steven Persell, MD, MPH; Abel Kho, MD; Nancy Dolan, MD; Muriel Jean-Jacques, MD

Organization: Northwestern University

Project Dates: 9/1/2008 – 8/31/2012 (including one-year no cost extension)

Federal Project Officer: Joy Basu, MD

Grant Award Number: 5R18HS017163-04

ABSTRACT

Purpose: The overall goal of this study was to determine whether a health information technology-enabled quality improvement (QI) strategy could improve performance on a set of 18 measures of quality of care for 4 chronic conditions and five preventive services.

Scope: One academic General Internal Medicine faculty practice.

Methods: We implemented a multifaceted QI intervention using electronic health record tools to improve quality measurement (including capture of contraindications and patient refusals), make point-of-care reminders more accurate, and provide more valid and responsive clinician feedback (including lists of patients not receiving essential medications). We used time series analysis to examine changes in quality of care during the following year. We subsequently added paper reminders for physicians to review prior to entering the examination room and analyzed whether this further improved quality.

Results: During the first year of the intervention, performance improved significantly for 14 measures. For 9 measures, the primary outcome improved more rapidly during the intervention year than during the prior year ($p<0.001$ for 8 measures, $p=0.02$ for 1). Adding paper reminders did not improve quality of care overall or for physicians whose performance lagged. The medical exceptions were almost always valid. Outreach to patients who refused services was ineffective.

Word count: 200

Key Words: health information technology, quality improvement, performance measurement

Purpose

The overall goal of this study was to determine whether a health information technology (HIT)-enabled quality improvement strategy could improve performance on a set of 18 measures of quality of care for 4 chronic conditions and five preventive services. The specific aims were:

Aim 1: Create simple, standard ways for clinicians to document *patient reasons* or *medical reasons* for why quality measures are not met. We will assess the use of these exception codes, the impact of exception reporting on measured levels of quality, and the impact of using these codes on physician satisfaction and self-reported efficiency.

Aim 2: Use the exception codes that clinicians enter (i.e., patient reasons and medical reasons for not providing a recommended therapy or preventive service) to target three strategies for quality improvement:

    a. Peer review of all medical reasons for not adhering to guidelines followed by academic detailing if a clinician enters an unjustified reason for not following guidelines
    b. Counseling for all patients whose physician enters an exception code stating that the patient cannot afford a needed medication to determine ways of overcoming barriers
    c. Educational outreach to all patients who refuse recommended interventions, including mailing of plain language health education materials or DVDs.

We will assess the frequency with which these interventions lead to changes in care.

Aim 3: Provide clinicians with highly accurate information on patients' quality deficits immediately prior to each patient's visit as part of routine work flow. We will assess whether this intervention increases a) provision of recommended therapies and tests, and b) documentation of exception codes.

Scope

<u>Background and Context</u>

Quality measurement techniques have vastly increased in sophistication during the past few decades and now allow for meaningful comparisons between healthcare facilities or health plans. However, as currently practiced, these methods cannot be used to raise healthcare quality to the highest possible level. Measures that depend on data collected for administrative purposes inevitably have measurement inaccuracies at the individual patient level. Patients may incorrectly be considered eligible for a measure; appear to fail a quality measure they have truly met because data satisfying the measure was not captured; or have reasons the measure was not appropriate for them (i.e., exclusion criteria) that the measurement system failed to detect. As a result of these limitations in the measurement systems, quality benchmarks are typically far less than 100%. "High performing" physician groups may only reach benchmarks in the 80 to 90% range. The gap between the benchmark and 100% is attributed to unidentified exceptions, patient preferences or measurement error. Thus, the true failure rate for recommended tests or therapies remains obscured. While these kinds of measurements permit valid comparisons—such as between health plans—imprecise measurement methods can never be used as the foundation upon which to build a healthcare system that strives to deliver near 100% high quality care for chronic disease care and prevention.

Information technology, especially electronic health records (EHRs), has the potential to revolutionize how quality is measured and how information is used to achieve truly outstanding levels of quality. However, even advanced EHRs have not incorporated quality measurement and have lacked seamless systems to use highly accurate data for comprehensive quality improvement. Most EHRs have the capability to deliver point of care (POC) alerts as quality improvement tools. However, POC alerts are often wrong because they do not capture

available exclusion criteria (e.g., the clinician receives a mammogram alert for a patient who had a bilateral mastectomy). In addition, alerts are often wrong because an exclusion criterion was never captured in an electronic format that the system could use (e.g., the patient has repeatedly refused pneumococcal vaccination, but this information is only in the doctors' notes). High rates of "false positive" alerts may contribute to clinicians' gradual inattention to alerts ("alert fatigue").

Quality measurement needs to be embedded within EHR systems and become more dynamic, accurate and detailed in order to provide the highest level of care possible to all patients. This project was designed to build on previous work done at the Northwestern Medical Faculty Foundation and create systems that improve our quality data and seamlessly link this data to practice-level quality improvement programs and POC interventions.

Setting
The main study took place at the Northwestern Medical Faculty Foundation's General Internal Medicine clinic. The clinic had been on the same EHR (Epic) for over ten years. During the two years prior to the study, physicians received printed quarterly reports of their performance on quality measures (including versions of 12 measures included in this intervention) without individual patient data. Interruptive (i.e., "pop-up") point-of-care reminders with links to order entry were active for many clinical topics but were rarely used. Some measures included limited medical exceptions (e.g., a documented drug allergy), but there was no mechanism for clinicians to record and capture other medical reasons and patient reasons for not following recommendations. These reminders were discontinued three months before the intervention began.

The dissemination sites for this study were four Family Medicine and Internal Medicine community practices in the Northshore University Health System, an integrated healthcare system north of Chicago. The practices had all been on Epic for several years.

Participants
All patients eligible for one or more quality measures cared for by attending physicians were included. At the start of the intervention, 12,288 patients were eligible for any measure. 75.7% were female, mean age was 53.1 years, 48.0% were white, 23.7% were black, 16.2% were Hispanic and 12.1% were of other or unknown race. 67.0% had commercial health insurance, 26.8% had Medicare, 3.4% had Medicaid and 2.8% were uninsured. There were 39 internists at the practice; 49% were female and 92% had worked at the practice for at least 2 years. The number of patients eligible for each quality measure ranged from 106 to 7,462.

Targeted Conditions and Services
The study originally targeted 18 measures of quality of care for 4 chronic conditions and five preventive services. Two of the measures were related to blood pressure control (patients with and without diabetes). They could not be implemented by the time of the study because of technical limitations, and they were therefore not included as part of the main reports. Table 1 shows the measures.

Table 1. Quality of Care Measures

Measure (Number of Eligible Patients) *	Denominator criteria	Numerator Satisfied	Exception Applied when Numerator not Met†
Coronary Heart Disease			
Antiplatelet drug (1202)	Coronary heart disease diagnosis	Antiplatelet drug on active medication list	Anticoagulant prescribed, medical reason, patient reason,
Lipid lowering drug (1202)	Coronary heart disease diagnosis	Lipid-lowering drug on active medication list	Medical reason, patient reason, LDL <100 mg/dl within the last 365 days
Beta blocker after MI (235)	Myocardial infarction diagnosis	Beta blocker on active medication list	Medical reason, patient reason, beta blocker allergy
ACE inhibitor or ARB (443) ‡	Coronary heart disease diagnosis and diabetes diagnosis	ACE inhibitor/ARB on active medication list	Medical reason, patient reason, ACE inhibitor and ARB allergy
Heart Failure			
ACE inhibitor or ARB in LVSD (276) ‡	Heart failure diagnosis	ACE inhibitor/ARB on active medication list	Medical reason, patient reason, ACE inhibitor and ARB allergy, LVEF >40%
Beta blocker in LVSD (276) ‡	Heart failure diagnosis	Beta blocker on active medication list	Medical reason, patient reason, beta blocker allergy, LVEF >40%
Anticoagulation in atrial fibrillation (106)	Heart failure and atrial fibrillation diagnosis	Anticoagulant on active medication list or referred to anticoagulation clinic	Medical reason, patient reason
Diabetes Mellitus			
HbA1c control (1814) ‡	Diabetes diagnosis	HbA1c < 8.0%	Medical reason, patient reason
LDL control (1595) ‡	Diabetes diagnosis, ≥ 50 years and female, or male	LDL-C < 100 mg/dL	Medical reason, patient reason , prescribed high potency statin§

Aspirin for primary prevention (1695)	Diabetes diagnosis, no diagnosis of coronary heart disease and age ≥40 years	Antiplatelet drug on active medication list	Medical reason, patient reason, aspirin allergy, anticoagulant prescribed
Nephropathy screening or management (1814) ‡	Diabetes diagnosis	Test for nephropathy in past year or ACE inhibitor or ARB on active medication list	Medical reason, patient reason

Prevention

Screening mammography (3539) ‡	Women 50 to 69 years	Mammography in past 2 years	Medical reason, patient reason
Cervical cancer screening (7462) ‡	Women 21 to 64 years	Cervical cytology in past 3 years	Medical reason, patient reason
Colorectal cancer screening (7067) ‡	Age 50 to 80 years	FOBT in 1 year, sigmoidoscopy in 5 years, colonoscopy in 10 years or DCBE in 5 years	Medical reason, patient reason
Pneumococcal vaccination (2966) ‡	Age ≥ 65 years	Pneumoccocal vaccine ever	Medical reason, patient reason
Osteoporosis screening or Therapy (1816) ‡	Women age ≥ 65 years	Bone density with central DXA after age 60 or medical therapy for osteoporosis¶	Medical reason, patient reason

ACE = angiotensin converting enzyme. ARB =angiotensin-receptor blocker. CHD = coronary heart disease. DCBE = double contrast barium enema. LVSD = left ventricular systolic dysfunction. LDL-C = low-density lipoprotein cholesterol.

*Number eligible February 1, 2008. Patients were eligible for quality measures if they had two or more office visits in the prior 18 months, No patients younger than 18 years of age were included.

† Exceptions that are medical reasons, patient reasons, erroneous diagnoses, or LVEF >40% are entered manually by clinicians. All other exceptions are measured automatically from existing coded data fields in the electronic health record.

‡ Indicates that point-of-care reminder was newly added at the start of the intervention. All others had reminders that were redesigned at the start of the intervention.

§ Prescribed atorvastatin 80 mg, rosuvastatin 40 mg, or simvastatin 80 mg tablet

¶ Prescribed a bisphosphonate, systemic estrogen, selective estrogen receptor modulators, parathyroid hormone, or calcitonin.

Methods
Study Design
The overall study design was a time-series analysis that examined changes in quality of care and changes in the rate of improvement in the performance measures between baseline and the initial implementation of the point of care alerts and the physician feedback system (phase 1, 12 months).

Data Sources/Collection
All data for calculating the performance measures was extracted from our electronic health record database and our enterprise data warehouse using structured query language.

Interventions
Phase 1: Interventions included point-of-care reminders, linked order sets, point-of-care tools within reminders for documenting exceptions (i.e., patient refusals, inability to afford medications, and contraindications or adverse reactions to recommended interventions), quarterly performance reports, and monthly lists for each physician in the practice of their patients who were not prescribed "essential" medications. In addition, there was a patient-focused intervention; if a patient refused a recommended procedure and the physician documented this, the patient was sent information about the benefits of the intervention (e.g., medication or preventive service) and contacted to see if s/he wanted to change his/her decision and receive the intervention.
Phase 2: In addition to the interventions described above, a list of unsatisfied quality measures was printed on a sheet for physicians to review prior to entering the examination room.

Measures
See Table 1.

Planning and Implementation of the Phase 1 Interventions
We modified existing point-of-care reminders or created new ones within the EHR for chronic disease (coronary heart disease, diabetes and heart failure) and prevention topics. We sought to have a fairly comprehensive set of reminders for common medical problems and preventive services for adult patients in a general medical practice. This would allow clinicians to use the reminder system routinely in their workflow rather than in only a few specific circumstances.

Instead of interruptive alerts, we used a minimally intrusive reminder: a single tab on the side of the EHR screen which was highlighted in yellow if any measure was not satisfied and an exception was not documented. (Supplemental Figures 1 and 2 in the Persell *Med Care* 2011 publication show the appearance of the reminders). We added standardized ways to capture patient reasons (e.g., refusals) or medical reasons that were exceptions for individual reminders (e.g., antiplatelet drug not prescribed in coronary heart disease due to a medical reason) within the reminder system of the EHR. Clinicians could also enter global exceptions for all quality measures (e.g., to indicate when a patient had a terminal disease) or indicate that a chronic disease previously recorded was not present within the preventive health tracking system of the EHR. These exceptions were then used to suppress multiple reminders and were included in the calculation of more than one quality measure when appropriate. We performed peer review of medical exceptions for the first 7 months of the intervention. Detailed methods and results of this process of have been reported. If no medical reason was identified or an inappropriate medical reason was present, we did not count these patients as having exceptions, and physicians were provided with patient-specific feedback. Preventive services performed elsewhere could all be recorded in standardized ways.

The UPQUAL intervention was implemented February 7, 2008. We held a one hour meeting and sent clinicians electronic training materials to encourage them to use the decision support tools and to record exceptions when present. We sought to create an expectation that for process of care included in this study, patients should either receive the recommended care or have the reason clearly stated why it was not. However, we emphasized that performance data would not be used to determine compensation during the study period. We informed clinicians that medical exceptions would be peer reviewed.

We gave physicians printed lists each month of their patients who appeared to be eligible for an indicated medication but were not receiving it and had no exception recorded. Quarterly performance reports were continued as before the start of the study. For each measure, we added the number of patients eligible, the number who satisfied it, and the number with exceptions.

Evaluation and Outcomes for Initial Implementation
At each time point, patients were eligible for a measure if they had 2 or more office visits in the preceding 18 months, were cared for by an attending physician, and met the other measure criteria (Table 1). To maximize the detection of quality problems, we included patients when ICD9-CM disease codes were recorded on the active problem list, past medical history, or as prior visit diagnoses. We used Structured Query Language to retrieve data from an enterprise data warehouse that contains data copied daily from the EHR. For each of the 25 months of the evaluation period, all patients were classified for each measure for which they were eligible as: a) satisfied, b) did not satisfy but had an exception, or c) did not satisfy and had no documented exception. The primary outcome for each measure was calculated as the number of patients who satisfied the measure divided by the total number of eligible patient excluding those with an exception. As an equation, the primary outcome = number satisfied / [number eligible – number not satisfied with an exception]. We also analyzed separately for each measure the proportion of eligible patients who satisfied the measure and the proportion of all eligible patients who did not satisfy the measure and had exceptions.

Statistical Analysis for Initial Implementation
We calculated the primary and secondary outcomes for each of the 16 performance measures for the first of each month from February 1, 2007 through February 1, 2009. This yielded three 25- point time series for each measure. A linear model was fit to each series using time as a continuous predictor, intervention as a dichotomous indicator variable, and a term for the interaction between time and intervention. Next, we determined the autoregressive order of the model residuals by minimizing Akaike's information criterion. Finally, we fit a linear regression model with autoregressive errors (using the appropriate number of autoregressive parameters, if any were necessary) to each series. These fitted models were used to tests statistical significance. To ensure model validity, we examined several residual diagnostics, the Jarque-Bera and the Shapiro-Wilk tests for normality of residuals, and normal Q-Q and autocorrelation plots.

To examine whether some physicians were able to achieve very high levels of care for drug prescribing process measures, we calculated the proportion of physicians who met the primary outcome for all (100%) or nearly all (90 to 99%) of their patients for each drug prescribing measure on February 1, 2008 and February 1, 2009 and analyzed differences between the two time points using Fisher's exact test. Analyses used SAS version 9.1 (SAS Institute Inc., Cary, NC) and R software package version 0.10-16 (R Foundation for Statistical Computing, Vienna, Austria).

Implementation of the Phase 2 Intervention
The nurses in the GIM clinic typically record vital signs and any comments for the physician (e.g., "needs medication refill") on a sheet that is left in a box outside the examination room. In February of 2009, we implemented a system that queried the EHR for outstanding quality deficits when the patient registered and printed these for the rooming nurses to use in lieu of their previous rooming sheets. All other quality measurement and feedback remained the same as in Phase 1.

Assessment of the Validity of Recorded Medical Exceptions
We performed peer review for all exceptions entered into the EHR beginning in February 2008, with the goal of performing at least 600 reviews. Every 1 to 2 weeks, we extracted all medical exceptions recorded in the EHR since the previous review. One physician reviewed medical records to collect the reason for the exception and additional clinical information needed to judge the validity of the exception. When the clinical reasoning was unclear, the peer reviewer would request clarification from the treating clinician. Three board-certified internists met regularly to review the exceptions. For some recorded medical exceptions, no real medical reason was noted (for example, "Cervical cancer screening was not done because I will do it at the next visit"). The internists judged these exceptions as having no medical reason present. In these cases, we notified physicians that they had used the exception reporting improperly, and we removed these exceptions from the clinical information system. The group reviewed the remaining medical exceptions and judged them as appropriate, inappropriate, or of uncertain appropriateness by consensus. When a consensus was not reached or the appropriateness was uncertain, 1 physician reviewed the medical literature, requested advice from specialists when needed, and the group discussed the case again until consensus was reached. Once the group had some experience with common appropriate medical exceptions, straight-forward, appropriate exceptions were classified as such after single-physician review. We gave feedback directly to the treating clinician in cases in which the medical exception was judged to be inappropriate and gave recommendations to change management. We provided information from the medical literature or expert opinion in cases that were judged as uncertain when the peer-review panel felt there was valuable information for the treating physician to consider. We recorded all time spent performing the peer-review process.

Evaluation of the Effect of Outreach to Patients Who Refused Recommended Services
During the outreach period, each week we performed an automated search of the EHR to identify patients with any new refusals recorded to electronic reminders for the five preventive services above. A non-clinician care manager performed the following tasks: manually reviewed the medical chart to determine if any specific barriers to obtaining the refused service were documented, mailed patients brief educational materials that included plain language educational brochures relevant to each topic, and attempted telephone contact. When telephone contact was successful, the care manager attempted to identify and resolve any barriers to obtaining the service by providing education, and when appropriate, obtaining needed referrals, facilitated the scheduling of necessary appointments, or referring the patient back to the practice clinicians if questions arose.

The outreach was stopped after one year because of a low rate of getting patients to accept services. We compared outcomes during Phase 1 to the subsequent year when outreach did not occur. We analyzed the time to reciept of the refused sevice for the outreach cohort and the post-outreach cohort using curves obtained employing the Kaplan Meier method. We used Cox regression models to calculate proportional hazards adjusted for patient characteristics (age, gender, race/ethnicity and insurance type when appropriate).

<u>Randomized Controlled Trial of Outreach to Patients Who Did Not Complete Colonoscopy</u>
Our analyses during phase 1 showed a high rate of patients not completing colonoscopy after it was ordered. We therefore added to our study a randomized trial of outreach to these patients. Using data contained in the EHR, we identified patients 50 to 79 years old who had received an order for a colonoscopy between November 2007 and January 2009 but had not received a colonoscopy in the 3 months following the order. For patients randomized after January 2009, we lowered the upper age limit to 75 years to coincide with the revised USPSTF screening recommendations. We identified 1036 patients with uncompleted colonoscopy orders. After excluding ineligible patients, 628 patients remained. These were stratified based on prior colorectal cancer screening (ever vs. never) and then randomly assigned in equal numbers to either the control or intervention arms. Patients in the intervention group were sent a mailing consisting of a personalized reminder letter from the physician, an educational brochure, and a digital video disk (DVD) about colorectal cancer and colorectal cancer screening. The tailored letter, which included their physician's digital signature, included the date the patients talked with their physician regarding colorectal cancer screening. It reminded patients that the physician had ordered a screening colonoscopy and reiterated the importance of screening even if patients had no symptoms or family history. The educational brochure provided information on colorectal cancer and options for screening. Both the letter and the brochure included the clinic telephone number and asked patients to call for an updated order. The DVD outlined common myths and questions regarding colorectal cancer and screening.
The primary study outcome was the completion of colorectal cancer screening using fecal occult blood testing, flexible sigmoidoscopy, or colonoscopy during the 3 months after randomization. We analyzed outcomes using time-to-event methods (i.e., Cox regression with the intervention group as the sole independent variable) according to the intention-to-treat principle.

<u>Dissemination to Four Suburban Primary Care Sites</u>
We applied the same principles to four coronary artery disease quality measures at four primary care sites in North suburban Chicago that share the same electronic health record. Components of the intervention were implemented sequentially. First, POC decision support that included simple exception recording was implemented in July 2008. Feedback reports, including lists of patients who did not satisfy the measure and who had no exceptions were begun in September 2009. This health system also announced in the fall of 2009 that it intended to apply financial incentives to these and other quality measures. We used time series analyses in a similar fashion to assess changes.

Results
The full results of this study cannot be described within the page limitations of this report. Therefore, only the most important results are presented. Full results are in the cited papers.

<u>Phase 1</u>
Table 2 shows the proportion of patients satisfying the primary outcome for each of the 16 fully implemented measures 1) one year before the start of the intervention (2/1/07), 2) at the start of the intervention (2/1/08), and 3) one year after the start of the intervention (2/1/09). In addition, Table 2 shows the rates of change for each measure during the year before (2/1/07 to 2/1/08) and the year after the intervention (2/1/08 to 2/1/09), as obtained from the fitted regression models.

During the year before the start of the intervention, performance improved significantly for eight measures, did not change for six, and declined for one (Table 2). Temporal trends could not be calculated for cervical cancer screening because exceptions were recorded without associated dates during the pre-intervention period. During the year after the start of the intervention

(2/1/08 to 2/1/09), performance improved significantly for 14 measures, improved non-significantly for another (hemoglobin A1c control; p = 0.08), and declined for one (screening mammography; Table 2). The final columns in Table 2 show the modeled difference in the rate of change in performance between the pre-intervention and the intervention years and the statistical significance of the difference.   During the intervention year, the rate of improvement in performance was significantly greater for 9 measures (p<0.001 for eight measures, p =0.02 for one measure) and of borderline significance for another (hemoglobin A1c control; p = 0.09). Another four measures improved during the post-intervention period, but the rates of improvement were similar to the pre-intervention period. The rate of improvement in performance for osteoporosis screening was lower during the intervention year than the pre-intervention year. The absolute rate of screening mammography declined.

The improvements in performance during the intervention year were due to a combination of more patients satisfying the measures (e.g., aspirin being prescribed for primary prevention for patients with diabetes; completion of colon cancer screening) and documentation of exceptions (e.g., medical or patient exception for anticoagulation for patients with heart failure and atrial fibrillation). For 9 of the 16 measures, the proportion of eligible patients satisfying the numerator criteria for the measures significantly increased, ranging from 0.8 to 9.5 percent increase per year. There was a significant 4.6 percent decline in completion of screening mammography and a 1.6 percent decline in patients with heart failure who were prescribed an angiotensin converting enzyme inhibitor or angiotensin receptor blocker, which was of borderline statistical significance (p=0.05). Physicians regularly used the tools to document exceptions. In the year after the start of the intervention, the modeled rate of change in the proportion of patients with recorded exceptions was significantly greater than zero for all measures.

Phase 2
Performance improved significantly for 8 of the 16 measures during Phase 2. For all of these measures, performance had improved significantly during Phase 1. Performance of screening mammography declined significantly during Phase 2; this was already declining in Phase 1, which we had attributed previously to a shortage of trained radiologists and prolonged waiting times at our institution. Performance decreased for two other measures during Phase 2: 1) prescription of anticoagulants for patients with atrial fibrillation and heart failure, and 2) nephropathy screening or management for patients with diabetes. Both of these had previously shown an improvement in performance during Phase 1. Performance did not change during Phase 2 for antiplatelet drug prescribing for patients with coronary artery disease; performance had increased during phase 1 and remained stable at a very high level (approximately 95%). Glycemic control (hemoglobin A1c < 8 mg/dl) did not change throughout the study.

We were particularly interested in whether physicians with the worst performance during Phase 1 improved during Phase 2. Some physicians said the point of care electronic reminders were not helpful because they did not immediately open patients' electronic records and may not have completed their charting until later. Having the reminder sheets outside the examination room should ameliorate this. However, we found no evidence that they improved disproportionately during Phase 2 (data not shown in this report; see published paper).

Table 2. Percentage of Patients meeting Quality Measures and Modeled Rates of Change for Quality Measures in the Year Before and the Year During the Intervention*

Measure	Percentage of Eligible Patients without Quality Deficit			Modeled rate of change, % per year†				Modeled Difference in Rate of Change Between 08-09 and 07-08†	
				Pre-Intervention		Post-Intervention			
	2/1/07	2/1/08	2/1/09	07-08	P	08-09	P	% per year	P
Coronary Heart Disease									
Antiplatelet drug	88.9	89.2	95.1	0.5	NS	**4.9**	**0.005**	**4.1**	**0.02**
Lipid lowering drug	88.4	87.1	93.3	**-1.5**	**<0.001**	**5.7**	**<0.001**	**7.3**	**<0.001**
Beta blocker in MI	85.8	89.6	94.0	**3.3**	**<0.001**	**5.8**	**<0.001**	1.9	NS
ACE inhibitor or ARB	82.7	84.4	89.7	**1.9**	**<0.001**	**5.1**	**<0.001**	**3.2**	**<0.001**
Heart Failure									
ACE inhibitor or ARB in LVSD	84.9	85.1	89.3	-0.1	NS	**5.2**	**<0.001**	**5.3**	**<0.001**
Beta blocker in LVSD	81.4	82.7	90.2	**1.6**	**0.02**	**7.3**	**<0.001**	**5.7**	**<0.001**
Anticoagulation in atrial fibrillation	66.7	65.1	85.3	-3.2	NS	**16.6**	**0.004**	**18.1**	**<0.001**
Diabetes Mellitus									
HbA1c control	65.3	64.7	66.0	-0.7	NS	2.1	NS (0.08)	3.1	NS (0.09)
LDL-C control	50.9	51.6	56.3	**0.8**	**0.04**	**4.8**	**<0.001**	**4.1**	**<0.001**
Aspirin for primary prevention	75.4	76.4	90.5	2.1	NS	**11.2**	**<0.001**	**8.8**	**<0.001**
Nephropathy screening or management	81.5	81.4	87.7	-0.2	NS	**5.7**	**<0.001**	**6.3**	**<0.001**
Prevention									
Screening mammography	79.3	80.8	77.1	**2.0**	**<0.001**	**-4.3**	**<0.001**	**-6.3**	**<0.001**
Cervical cancer screening	‡	83.8	86.2	‡	‡	**2.4**	**<0.001**	‡	‡
Colorectal cancer screening	53.7	56.6	62.0	**2.8**	**0.007**	**4.9**	**<0.001**	2.2	NS
Pneumococcal vaccination	72.3	80.1	89.9	**8.5**	**<0.001**	**9.1**	**<0.001**	0.2	NS
Osteoporosis screening or Therapy	72.3	77.9	82.0	**5.7**	**<0.001**	**4.1**	**<0.001**	**-1.7**	**0.02**

* ACE = angiotensin-converting-enzyme; ARB = angiotensin receptor blocker; LDL-C = low density lipoprotein cholesterol; LVSD = left ventricular systolic dysfunction; MI = myocardial infarction. Quality measures were calculated as the number who satisfied the measure / (number eligible – number not satisfied with an exception). † Rates of change were derived from the linear regression models with autoregressive errors when necessary as described in the methods. ‡ Performance for the cervical cancer screening measure could not be calculated accurately prior to January 2008 because of missing date information for exceptions.

Validity of Medical Exceptions

A total of 87 physicians (49 resident and 38 attending) recorded 650 medical exceptions from February to September 2008. Physicians used the medical-exception–reporting tool 36 times (5.5% [95% CI, 3.9% to 7.6%]) when the reason for not following the decision support was not due to any medical reason. Of the remaining 614 medical exceptions, 93.6% (CI, 91.4% to 95.4%) were judged as appropriate, 3.1% (CI, 1.9% to 4.8%) inappropriate, and 3.3% (CI, 2.0% to 5.0%) of uncertain appropriateness. Frequencies of inappropriate and uncertain exceptions were 7 (6.9% [CI, 2.8% to 13.6%]) and 10 (9.8% [CI, 4.8% to 17.3%]) for coronary heart disease; 0 ([CI, 0.0% to 4.3%]) and 2 (2.4% [CI, 0.3% to 8.4%]) for heart failure; and 10 (10.8% [CI, 5.3% to 18.9%]) and 8 (8.6% [CI, 3.8% to 16.2%]) for diabetes. For preventive service, nearly all medical exceptions were judged appropriate: 334 (99.4% [CI, 97.9% to 99.9%]). Only 2 (0.6% [CI, 0.1% to 2.1%]) were inappropriate, and none was of uncertain appropriateness. Of all medical exceptions recorded by physicians, 78 (12.7% [CI, 10.2% to 15.6%]) were instances in which a clinician recorded that a diagnosis that triggered a quality alert was not present. Peer reviewers disagreed with these exceptions 10.2% (CI, 4.5% to 19.2%) of the time and were uncertain 2.6% (CI, 0.3% to 9.0%) of the time.

Effect of Outreach to Patients Who Refused Recommended Services

In the outreach cohort, 407 patients had refusals documented for 520 preventive services. The corresponding numbers in the post-outreach cohort were 378 and 510, respectively. The outreach cohort was slightly older than the post outreach cohort, was more likely to have Medicare insurance and was more likely to have refused colorectal cancer screening or pneumococcal vaccination compared to the post-outreach cohort.

Few patients received any preventive service within 6 months of when the refusal was recorded in the EHR. There was no difference between the outreach cohort and the post-outreach cohort in the receipt of any refused preventive service, 6.1% in the outreach cohort, 4.8% in the post-outreach cohort (adjusted HR 1.3, 95% [CI 0.7%-2.5%]) (Table 3). Table 3 provides the number and percentages of patients in the two cohorts who refused services and rates of receiving each service within 6 months. There was variation from service to service in the difference between outreach and post-outreach cohorts (Table 3). However, the number of patients who received any individual service was generally small and the estimates of the impact of outreach for individual services are imprecise. Patient refusals were rarely documented for chronic disease services (only 39 occurred in the first cohort). Since this number was so small, we did not compare the outreach and post-outreach cohorts for chronic disease services or analyze these results further. Conducting outreach for the 407 patients with documented refusal of one or more preventive service required 214 hours of care manager time which included chart reviews, preparation of mailings, attempted and completed phone calls, and appointment coordination.

Table 3. Preventive service refused and subsequently received within 6 months.

Preventive service	Outreach n / N (%)	No Outreach n / N (%)	Adjusted hazards ratio (95% CI)
Patient with any refused service	25 / 407 (6.1)	18 / 378 (4.8)	1.3 (0.7 – 2.5)*
By service			
Colorectal Cancer Screening	11 / 249 (4.4)	5 / 191 (2.6)	1.7 (0.6 – 4.9)*
Breast Cancer Screening	3 / 89 (3.4)	5 / 118 (4.2)	0.9 (0.2 – 3.9)†
Cervical Cancer Screening	8 / 60 (13.3)	6 / 83 (7.2)	1.9 (0.6 – 5.4)†
Osteoporosis Screening	1 / 29 (3.5)	3 / 43 (7.0)	0.5 (0.04 – 4.6)‡
Pneumococcal Vaccination	2 / 93 (2.2)	2 / 75 (2.7)	0.6 (0.1 – 4.5)§

* Adjusted for age, gender, race/ethnicity and type of health insurance.
† Adjusted for age, race/ethnicity and type of health insurance.
‡ Adjusted for age and race/ethnicity.
§ Adjusted for age, gender and race/ethnicity.

Effect of Outreach to Patients Who Did Not Complete Colonoscopy

A total of 628 patients were randomized to the intervention (N=314) and control groups (N=314). The mean age of participants was 58.0 years (SD = 6.9); 92.7 % had no prior colorectal cancer screening. After 3 months, 9.9% of patients in the intervention group and 3.2% in the control group had completed colorectal cancer screening (hazards ratio 3.2, 95% confidence interval [CI] 1.6 – 6.5; p = 0.0014). Figure 1 displays the Kaplan-Meier curves for the primary endpoint. At six months following randomization, intervention participants remained significantly more likely to have completed colorectal cancer screening than control participants, but the effect was diminished (18.2% vs. 12.1%, hazards ratio 1.60, 95% [CI 1.1% to 2.4%]; p = 0.026). Although the effect of the intervention was sustained out to six months post randomization, there was no increase in the difference in completion rates between the intervention group between 3 and 6 months.

We reached 109 of the 314 intervention group participants (34.7%) by telephone to complete a brief process evaluation. Of those contacted, 95% reported receipt of the mailed intervention materials. Of those who acknowledged having received the mailing, 98% reported that they read the letter and the enclosed brochure and 30% reported watching the DVD.

Figure 1. Time to Screening Completion
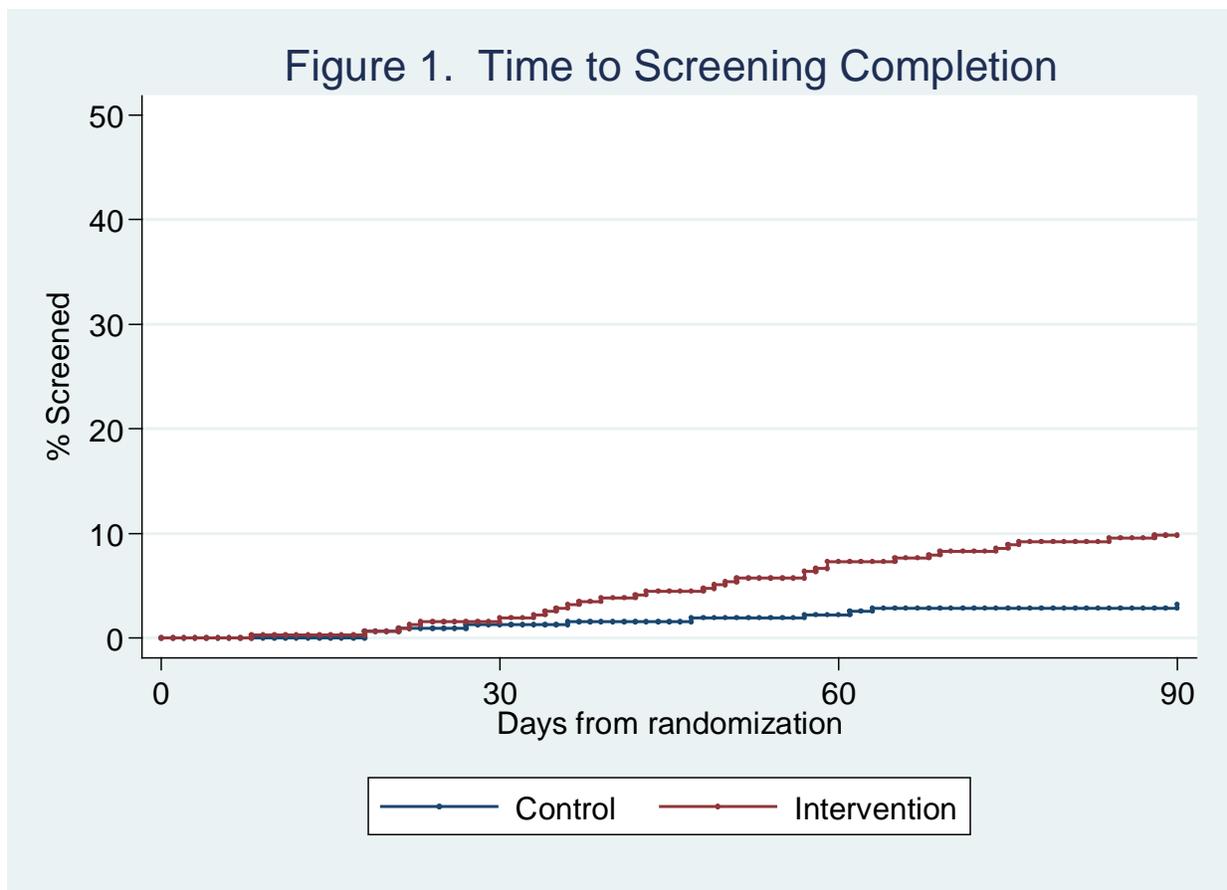
<u>Dissemination to Suburban Sites</u>
Performance improved for one measure following the implementation of POC reminders (antiplatelet drug for CAD) and improved for all 4 measures during the second intervention period. The rate of improvement in actual antiplatelet drug prescribing increased during the second intervention period. For the 3 other measures, improved performance during the second intervention period was entirely due to the recording of exceptions, and the rates of actual drug prescribing did not increase. Physicians recorded few exceptions to performance measures during the EHR alert intervention alone. Their recording of exceptions increased greatly during the second time period. Most exceptions were judged to be appropriate by peer review.

<u>Conclusions, Significance, and Implications</u>
There are a number of important lessons learned from the UPQUAL study.

A comprehensive, multi-faceted quality improvement approach can achieve sustained improvements in quality of care for multiple targets simultaneously.

Rather than focusing on a single condition or preventive service type (e.g., cancer screening), we choose a large, diverse set of targets. This allowed us to establish a usual work flow for all clinical decision support tools and quality improvement goals. Most, but not all, of the physicians in the practice embraced these and were able to take their quality of care to very high levels. We believe that this high rate of acceptance was possible because when the quality improvement tools were used properly, they actually had the potential to improve efficiency. All outstanding issues were easily visible with one click on a passive alert that was highlighted in yellow when there was a measure that had not been met. The alerts had key information included in them (e.g., date of last test and result, if applicable) and links that would allow clinicians to easily jump to other portions of the EHR to view information necessary to determine the appropriate action (e.g., jump to Health Maintenance plans, or jump to the Medication list). Most clinical decision support tools had linked order sets that allowed physicians to do place orders, enter diagnoses, and link orders to encounter diagnoses with a very limited number of clicks. In addition, physicians received timely, accurate feedback on their performance that allowed them to see the changes when they responded to point of care alerts. The feedback included exceptions entered as well as their overall performance on a measure. In addition, responding to an alert by entering an exception would suppress the alert for a year or more, allowing clinicians to avoid "alert fatigue." It is likely that these elements of the intervention had synergistic salutary effects on adoption and persistent use of the clinical decision supports and linked order sets.

Providing clinicians with lists of patients who need essential medications is a very strong lever for engaging clinicians in quality improvement activities.

We cannot separate out the independent effects of the different quality improvement activities. However, anecdotally, providing clinicians with a list every month of patients who were not receiving essential medications (e.g., lipid lowering therapy for a patient with coronary artery disease) seemed to be a more powerful motivator than simply providing them with a performance report (i.e., the percent of their patients with coronary artery disease who were prescribed lipid lowering therapy). This is not surprising because this information is immediately actionable by reviewing patients' charts and either contacting the patient to initiate therapy (or arranging for an appointment to discuss this) or documenting exceptions.

Paper reminders do not appear to improve quality of care beyond state-of-the-art electronic reminders.

Phase 2 of our study showed that paper reminders have little or no marginal value when added to electronic reminders. This is important because implementing systems to generate paper reminders requires both IT staff support and, for some sites, work flow changes. The time and effort to do this may be better spent doing additional academic detailing and training of providers whose performance continues to lag.

However, this finding may not be generalizable to practices that are less experienced users of electronic health record systems or to practices that fully embrace team-based approaches to quality improvement where nurses are empowered to act on reminders for preventive services (i.e., enter orders for clinicians to sign).

Physicians will routinely enter exceptions using point-of-care tools, and the medical exceptions entered are valid the vast majority of the time.

Within the context of our overall quality improvement program in which physicians received regular feedback about their performance, physicians were willing to regularly enter patient reasons (i.e., refusals, inability to afford services) and medical reasons (i.e., adverse events, relative contraindications). Moreover, in two different clinical settings, one in which there was no pay-for-performance program in place and another in which the intention to use pay-for-performance had been announced, the medical reasons were almost always valid. This is encouraging because most organizations do not have the resources to review all exceptions entered.

Nevertheless, we believe it is important for clinicians to be required to review cases of *inappropriate* exceptions when they are learning how to use these tools. For example, all clinicians should know that diabetic retinopathy is not a valid reason for not prescribing antiplatelet therapy to someone with comorbid coronary artery disease. Even though antiplatelet therapy does increase the risk of retinal hemorrhage, the benefits outweigh the risk. In our practice, we now have all new physicians review a set of cases for which the entered medical exception was deemed inappropriate based on literature review and/or expert opinion. Practices may want to monitor a sample of cases to determine the appropriateness of exceptions early on during implementation to ensure validity.

Some experts would like to create clinical decision supports with exception reporting tools that allow physicians to select a medical reason from a list (e.g., renal insufficiency for an ACE inhibitor, or bradycardia for a beta blocker for heart failure). However, in our experience, most of these medical reasons are _relative_ contraindications; the majority of patients may actually be able to tolerate the recommended therapy and achieve therapeutic benefits if managed carefully. Providing a drop-down list of medical reasons runs the risk of implying that these are _absolute_ contraindications, which could have the unintended consequences of decreasing the number of people who actually receive the recommended service.

Outreach to patients who have refused a recommended service is not effective at convincing patients to change their mind and receive the service.

We thought that sometimes providers may not have time to fully explain the benefits of preventive services or may not explain them in a way that overcomes literacy or cultural barriers; providing patients with additional information using state-of-the-art health communication tools could reverse their decision. However, this was not the case. In discussions with the physicians in our practice, most were reluctant to enter that the patient refused a service until the patient gave a strong, even vehement, refusal. In this setting,

outreach was futile. The results could be different if physicians had a lower threshold for entering patient refusals. This could be especially important for practices with financial incentive programs for quality or pay-for-performance programs; in these situations, physicians may be less inclined to have multiple discussions with patients about their need for services.

Electronic health record systems can be used for targeted outreach to patients who fail to complete recommended screening tests (e.g., colonoscopy). However, more effective interventions are needed to make this worthwhile on a large scale.

Our results were somewhat more encouraging for outreach to patients who failed to complete a colonoscopy. However, the absolute increase in colonoscopy completion was small. Thus, although we showed that it is possible to accurately use EHR systems to target patients for outreach, more effective interventions are needed to motivate patients to complete this test.

Once established, this type of quality improvement methodology can be sustained at minimal cost.

Creating the point-of-care reminder systems and the quality measurement programs require significant time and effort to develop. However, once established, maintaining the alerts requires a fairly small effort to update the lists of medications or diagnostic tests relevant for the clinical decision support tools and quality measures. Generating the quarterly performance reports and the monthly essential medication lists also requires minimal time and effort, and we hope to reduce this further with web-based reporting tools that are accessible through the EHR home page. Two years after completion of the study, the UPQUAL quality measurement and improvement tools are fully operational, quality of care has generally remained stable or improved, and we are working to expand the system to include additional targets.

Publications

Main Study Aims

Persell SD, Kaiser D, Dolan NC, et al. Changes in performance after implementation of a multifaceted electronic-health-record-based quality improvement system. Med Care. 2011 Feb;49(2):117-25. PubMed PMID: 21178789.

Baker DW, Persell SD, Kho AN, et al. The marginal value of pre-visit paper reminders when added to a multifaceted electronic health record based quality improvement system. J Am Med Inform Assoc. 2011 Nov 1;18(6):805-11. Epub 2011 Jun 9. PubMed PMID: 21659446; PubMed Central PMCID: PMC3197995.

Persell SD, Dolan NC, Friesema EM, Thompson JA, et al. Frequency of inappropriate medical exceptions to quality measures. Ann Intern Med. 2010 Feb 16;152(4):225-31. PubMed PMID: 20157137.

Persell SD, Friesema EM, Dolan NC, et al. Effects of standardized outreach for patients refusing preventive services: a quasi-experimental quality improvement study. Am J Manag Care. 2011 Jul 1;17(7):e249-54. PubMed PMID: 21819171.

Cameron KA, Persell SD, Brown T, Thompson J, Baker DW. Patient outreach to promote colorectal cancer screening among patients with an expired order for colonoscopy: a randomized controlled trial. Arch Intern Med. 2011 Apr 11;171(7):642-6. Epub 2010 Dec 13. PubMed PMID: 21149742.

Persell SD, Khandekar J, Gavagan T, et al. Sequential Implementation of Electronic-Health-Record-Based Quality Improvement Strategies to Address Outpatient Coronary Artery Disease Care. In preparation.

Secondary Analyses to Assess Disparities and Changes in Racial/Ethnic Disparities

Jean-Jacques M, Persell SD, Hasnain-Wynia R, et al. The Implications of Using Adjusted Versus Unadjusted Methods to Measure Health Care Disparities at the Practice Level. Am J Med Qual. 2011 May 23. [Epub ahead of print] PubMed PMID: 21609941.

Jean-Jacques M, Persell SD, Thompson JA, et al. Changes in Disparities Following the Implementation of a Health Information Technology-Supported Quality Improvement Initiative. J Gen Intern Med. 2011 Sep 3. [Epub ahead of print] PubMed PMID: 21892661.

AHRQ Health Care Innovations Exchange
http://innovations.ahrq.gov/content.aspx?id=3122