HSR&D
Health Services Research & Development Service

FORUM
Translating research into quality health care for Veterans

October 2012

Commentary

Quality Measurement and the Ubiquitous Electronic Health Record

Joseph Francis, M.D., M.P.H., Office of Informatics and Analytics, VA Central Office, Washington, D.C.

Contents

A publication of the VA Office of Research & Development, Health Services Research & Development Service, Center for Information Dissemination and Education Resources, in conjunction with AcademyHealth. For past issues of FORUM, please visit www.hsrd.research.va.gov/publications/forum.

The Veterans Health Administration was an early adopter of quality measurement, performance accountability, and the use of electronic health records (EHRs) in the 1990s. These steps were seen as important to uniformly implementing evidence-based practices in prevention and chronic disease care across our entire delivery system. Largely as a result of the influence of quality measurement and EHRs on clinician behavior, VA achieved "best care anywhere" levels of performance on the Joint Commission's Core Hospital Quality Measures and the National Committee for Quality Assurance (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS).

Constraints of EHRs

Now that VA is well into its second decade of using quality measures and EHRs to guide its care of Veterans, the constraints of this approach are becoming more apparent. Most nationally-recognized quality measures were implemented in a "one size fits all" manner—appropriate perhaps when performance levels were under 50 percent, but not at current levels, which exceed 85 percent for most measures. Both qualitative and quantitative studies are now raising concern that trying to improve performance on our existing measures might direct clinical resources toward interventions that have either low benefit for Veterans, or the potential to cause more harm than good.[1,2] Since most of our quality measures focus on a single episode of care, they may also fail to capture appropriate clinical decisions and changes in the patient's health status and risk over a longitudinal timeframe. And, given that many Veterans obtain at least part of their medical care elsewhere,[3] can we ever say with confidence that VA care is a good value for the nation?

Our electronic health record, VistA-CPRS, remains limited in its ability to capture clinical concepts using standardized data elements. As a result, VA currently uses chart abstraction on a sample of Veterans to estimate performance on HEDIS and Joint Commission measures. Not only does this cost more than $12 million annually, it causes delayed and inadequate clinical feedback to treatment teams at the point of care. When structured clinical data is captured electronically, it often occurs through relatively inflexible clinical reminders, which can create challenges for harried clinicians due to the poor context-sensitivity of the reminders and their interference with workflow.

These limitations might have mattered little when VA was among the few large U.S. health systems using the EHR and had unmatched performance statistics. That reality will soon change. Nearly $20 billion of incentive payments have been made available to eligible hospitals and providers to accelerate EHR adoption as part of the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009. Providers and hospitals that adopt certified EHRs must demonstrate "Meaningful Use" by, among other requirements, generating and reporting electronic quality measures and public health information, and implementing clinical decision support systems. Although many commercially available EHRs currently share many

Director's Letter

In the late 1990s, the VA was ahead of most health care organizations in incorporating a robust set of quality measures as part of sweeping reforms. Some 15 years later, the VA is once again leading—this time in applying health services research to document the limitations of many widely used measures and to develop better patient-centered measures of quality. What has changed? One problem has been the sheer proliferation of measures, which can overwhelm busy clinicians. The second is that the VA of today is a much better performing system: when there is a lot of room for improvement, even imperfect measures can motivate change without risking inappropriate care. Finally, implementation of measures has shifted over time from an initial focus on using them to motivate innovation and local solutions to improve quality—"making the right thing the easy thing to do"—to using measures to enforce compliance with national guidelines. As a recent editorial notes, this change risks "undermining the VA's culture of continuous improvement."[1] The good news is that VA leadership recognizes and is addressing these issues. Even as the new Office of Informatics and Analytics expands our ability to measure practice in real-time for improvement purposes, VA is revising the "one-size" approach to measurement for accountability.

There is plenty of work for researchers in helping create better measures, but it is equally important to learn the best ways to implement them. Should measures be used to motivate individuals or practices? How should incentives be designed when care is the responsibility not of an individual clinician but of a patient-aligned care team? Should we "retire" measures when performance reaches a desired level? Answering these and other questions is an essential task if VA is to continue to meet the IOM's recent vision of the learning health care system.[2]

David Atkins, M.D., M.P.H., Acting Director, HSR&D

1. Kizer K.W. and S.R. Kirsh. "The Double-edged Sword of Performance Measurement," *Journal of General Internal Medicine* 2012; 27:395-7.
2. National Research Council. 2012. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: The National Academies Press.

of the same constraints as VistA, HITECH is expected to stimulate considerable innovation to address these shortcomings.

Stage 2 and 3 of Meaningful Use will also include a demonstration of the ability to electronically send and receive information across providers using nationally recognized data standards. Other federally-supported initiatives, such as the development of the Nationwide Health Information Network (NwHIN), VA's own Blue Button technology (http://bluebuttondata.org), the population health tracking tools developed by Indian Health Service (http://www.ihs.gov/CIO/ca/icare/index.cfm), and the joint VA–Department of Defense commitment to develop an open-source integrated EHR (iEHR), offer a vision in which health information is shared seamlessly across multiple health systems and providers. Finally, rapid advances in computer science, especially the use of natural language processing for complex analytics (e.g., IBM's Watson system) are allowing use of much richer information to provide context-sensitive, patient-centric decision support.

New Ways to Measure Quality

The future in which the EHR is ubiquitous and health data is exchanged across providers and systems in real time allows news ways to measure quality, such as:

1. *Longitudinal quality measurement incorporating clinical actions*: assessing the adequacy of clinical care not just by the presence/absence of a given finding or intervention, but by the degree to which clinical actions over the course of time are consistent with scientific evidence and patient wishes.
2. *Risk-tailored quality measurement*: using the predicted risk of poor outcome to inform appropriate interventions. Risk is determined from the totality of available health data and modeled from populations that are relevant to the specific patient and situation.
3. *Patient-centered quality measurement:* allowing patients, and their loved ones, to identify realistic and desired health outcomes, expressed in terms that reflect personal values and goals for well-being and functioning in the community.

Any of these scenarios represents a major shift for quality measurement in VA and the nation. Several are already being tested and deployed—for example, we are implementing clinically-appropriate action measures for diabetes and a predictive model for mortality and hospital admission within primary care (albeit, at present, only with data obtained from VistA). Expanding this work will require deep and ongoing collaboration with informaticians and health services researchers who can provide the technical expertise and scientific objectivity to pilot, refine, and evaluate the measures themselves. Core challenges include developing means of capturing appropriate information, such as patient preferences, without interrupting clinical workflow, and establishing the psychometric properties of metrics that are based on personalized goals. The Office of Analytics and Business Intelligence looks forward to these partnerships.

References

1. Powell, A.A. et al. "Unintended Consequences of Implementing a National Performance Measurement System into Local Practice," *Journal of General Internal Medicine* 2012; 27:405-12.
2. Kerr, E.A. et al. "Monitoring Performance for Blood Pressure Management among Patients with Diabetes Mellitus: Too Much of a Good Thing?" *Archives of Internal Medicine* 2012; 172:938-45.
3. Trivedi, A.N. et al. "Duplicate Federal Payments for Dual Enrollees in Medicare Advantage Plans and the Veterans Affairs Healthcare System," *Journal of the American Medical Association* 2012; 308:67-72.

Response to Commentary

Moving toward a Patient-centered Performance Management System

Eve A. Kerr, M.D., M.P.H., and Rodney A. Hayward, M.D., both with the VA Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, Michigan

As Dr. Francis described in his commentary article, initial performance measures launched in VA and elsewhere nearly two decades ago helped drive quality improvement at a time when performance was universally low.[1] Presently, VA has achieved higher levels of performance on these "one size fits all" measures than most other health systems. As we and others have documented (see references 1 and 2 from Dr. Francis' commentary), when performance on such measures is high, aiming for even higher performance can push clinicians and facilities to overtreatment; such efforts may also ignore legitimate exceptions to the measures.[2] VA has the opportunity to move beyond those first step measures and develop a performance management system to help guide optimal patient-centered care—a system that takes into account the risks and benefits of tests and treatments for individual Veterans, as well as Veterans' own preferences.[3]

We propose that to move to such a system, we need to consider the net value in care provided—including expected benefits and risks at the *individual* patient level—and to elicit individual patient preferences for care. By net value, we mean the benefits or gains that can be expected from receiving care, minus costs and potential harms from the care. A patient-centered performance management system should define a point at which the net value of care for an individual is so high that it should almost always be provided, and not doing so would indicate poor care; and a point at which net value is so low (or even negative, in the instance of harmful care) that provision of that care should be an indicator of lower performance. In between the points of high-value and very-low value care lies a gray zone; in this gray zone, provision of care cannot be determined solely by the evidence and should instead be defined through informed patient-clinician decision-making. Documenting both that an informed discussion took place and what the patient's stated care preferences are should constitute high performance in the gray zone.

How would a patient-centered performance management system work? Consider a common scenario. Mr. John L., a 48-year old non-smoker, with treated hypertension and hyperlipidemia, but no history of cardiovascular (CV) disease presents to clinic with a systolic blood pressure of 144/82 mm Hg. How should the provider decide whether Mr. L needs to have his blood pressure medications adjusted? Currently, because the patient does not meet the "all or nothing" blood pressure < 140/90 measure, his provider would likely add a blood pressure medication to his regimen. In a patient-centered performance management system, the electronic health record (EHR) would automatically calculate, using a Veteran-specific risk model, Mr. L's 10-year risk for CV disease (about 5 percent). Then, the system would present the CV risk reduction (benefit) and side effects of specific medications or other treatments in a way that both provider and patient could understand; the system would also suggest whether the magnitude of the proposed treatment effects was large (high net value), moderate or small (moderate to low net value), or even harmful (negative net value). The EHR would also promote documentation of Mr. L's goals and preferences, such as avoiding side effects, having high levels of energy, and minimizing co-pays. Because the clinical impact from additional treatments in Mr. L's case is low, the provider and patient could decide together not to add an additional antihypertensive medication. The EHR would then prompt for documentation that the patient received information on the risk and benefit of treatment options and chose not to pursue treatment, thus fulfilling the performance measure for hypertension management. In this way, the system would not only guide real-time and patient-centered treatment, but also produce retrospective performance measures that reflect, for example, the efficiency of care (frequency of provision of high value care minus frequency of provision of very low value care) and patient-centeredness of care (documentation of patient preferences in moderate value care). Importantly, it would also incorporate the three elements Dr. Francis suggests, and we agree, are needed to advance performance measurement: risk-tailoring, capturing clinical actions, and incorporating patient preferences and goals.

Given the advances in the last decade in health services research, decision science, implementation research, and informatics, progressing to a patient-centered performance management system is within our grasp, particularly in VA. Success, however, will require committed partnerships between clinical, policy, and operations leaders, informaticians, and health services researchers. Indeed, health services research can help significantly impact VA care delivery by: developing approaches for defining high value, low value, and preference sensitive care; developing and testing use of VA-specific risk prediction models; developing decision support tools that incorporate individual risk and benefit information and testing alternate ways to present this information to patients and providers; conceiving and evaluating new measures to assess patient-centered quality of care, including measures of under- and over-treatment; and assessing how incorporation of patient-centered performance management influences patient experiences, prescribing, time involved in care provision, costs, and provider satisfaction.

References

1. Asch, S.M. et al. "Comparison of Quality of Care for Patients in the Veterans Health Administration and Patients in a National Sample," *Annals of Internal Medicine* 2004; 141(12):938-45.
2. Kerr, E.A. et al. "Building a Better Quality Measure: Are Some Patients with Poor Intermediate Outcomes Really Getting Good Quality Care?" *Medical Care* 2003; 41(10):1173-82.
3. Sussman, J.B. et al. "Individual and Population Benefits of Daily Aspirin Therapy: A Proposal for Personalizing National Guidelines," *Circulation: Cardiovasc Quality and Outcomes* 2011 May; 4(3):268-75.

Research Highlight

Validating the Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicators: How Well Do They Identify True Safety Events?

Amy K. Rosen, Ph.D., Center for Organization, Leadership, and Management Research (COLMR), VA Boston Healthcare System, Boston, Massachusetts

Ensuring patient safety, "freedom from accidental injury caused by medical care," is a high priority for the VA. A formal patient safety program was launched in 1997, promulgating the VA as a national leader in patient safety. This program included the establishment of the National Center for Patient Safety, developed to advance patient safety measurement and quality improvement (www.patientsafety.gov).

Despite the numerous patient safety initiatives underway, the VA does not yet have a national system in place for tracking patient safety events. The Patient Safety Indicators (PSIs), developed by the Agency for Healthcare Research and Quality (AHRQ) and released in 2003, represent a substantial contribution to the scientific detection of patient safety events. The PSIs are an evidence-based method designed to screen for potentially preventable adverse events that occur in the inpatient setting, such as complications following surgeries, other procedures, and some medical care. The PSIs are based on ICD-9-CM diagnosis and procedure codes obtained from administrative discharge data, making them readily available, cost-efficient, and easy to use. However, because they are based on administrative data (known for its coding variability), concerns related to their validity and reliability as patient safety "outcomes" are prevalent. Moreover, these concerns have heightened with increasing emphasis on the use of the PSIs for public reporting and pay-for-performance, rather than for how they were originally intended, for quality improvement and case-finding activities. The Centers for Medicare and Medicaid Services (CMS) currently posts six individual PSIs and a PSI composite on their Hospital Compare website and will be tracking these through their annual payment program (www.cms.gov/HospitalQualityInits/11_HospitalCompare.asp.). It is only a matter of time before the VA follows suit.

Recognizing the need to understand whether the PSIs identify true events, the VA HSR&D Service funded a study to examine the validity of the PSIs: "Validating the Patient Safety Indicators in the VA: A Multi-Faceted Approach" (SDR-07-002). One important component of this four-year study (2003-2007) was to assess the criterion validity of the PSIs (i.e., do cases flagged by the PSI algorithm represent true events based on medical record review, "the gold standard"?). We selected 12 of the 20 hospital-level PSIs for study based on their relevance to the VA population, observed VA rates, and their potential preventability (see Table).

Our 28 hospitals were drawn from a nationally representative sample of VA acute-care hospitals, selected based on individual PSI counts, PSI composite rates, and geographic distribution. From each hospital, we randomly selected four PSI-flagged medical records for medical record abstraction (112 cases per PSI). Two nurse-abstractors reviewed electronic medical records (EMRs) using standardized data abstraction instruments and guidelines based on AHRQ-developed tools or developed *de novo*. The EMR was reviewed for the occurrence of a safety-related event; patient clinical characteristics; clinical processes related to the event; and patient outcomes. We examined inter-rater reliability between the nurses, setting a target of 90 percent agreement.

PSIs Examined by Study

PSI #3	Pressure Ulcer
PSI #5	Foreign Body Left in During Procedure
PSI #6	Iatrogenic Pneumothorax
PSI #7	Central Venous Catheter-related Bloodstream Infections
PSI #8	Postoperative Hip Fracture
PSI #9	Postoperative Hemorrhage or Hematoma
PSI #10	Postoperative Physiologic and Metabolic Derangements
PSI #11	Postoperative Respiratory Failure
PSI #12	Postoperative Pulmonary Embolism or Deep Vein Thrombosis
PSI #13	Postoperative Sepsis
PSI #14	Postoperative Wound Dehiscence
PSI #15	Accidental Puncture or Laceration

To assess criterion validity, we calculated the positive predictive validity (PPV) of each PSI and associated 95 percent Confidence Intervals (CIs). PPV was calculated by dividing the number of true positives (TPs) by the number of flagged cases. We also examined false positives (FPs) to determine why they were flagged and how the PSI might be improved.

PPVs for selected PSIs varied considerably, ranging from a low of 28 percent for Postoperative Hip Fracture to a high of 87 percent for Postoperative Wound Dehiscence. PPVs for the other PSIs were relatively moderate, ranging from 43 percent for PE/DVT to 75 percent for Postoperative Hemorrhage/Hematoma. Of the nine surgical PSIs, seven PPVs were >60 percent, demonstrating better predictive validity than the medical PSIs. A common reason for false positives included conditions that were present-on-admission. For example, among false positive cases of Pressure Ulcer, 83 percent were present-on-admission. Another primary cause for false positives was misidentification of non-elective admissions as elective.[1]

Given the relatively moderate PPVs found, we recommend that the PSIs should continue to be used primarily for quality improvement and case-finding activities. The possible exceptions to this are some of the surgical PSIs (those with the highest PPVs); however, even these should be used only for public reporting, given that coding revisions are needed to improve most of the indicators. Nonetheless, the PSIs are a step in the right direction; it is important to understand what the indicators detect and where the greatest opportunities for quality improvement and case finding lie.

References

1. Rosen, A.K. et al. "Validating the Patient Safety Indicators (PSIs) in the Veterans Health Administration: Do They Accurately Identify True Safety Events?" *Medical Care* 2012; 50(1):74-85.

Research Highlight

Performance Measure for Heart Failure Care

Paul Heidenreich, M.D., M.S., Center for Healthcare Evaluation, Chronic Heart Failure QUERI, VA Palo Alto Healthcare System, Palo Alto, California

The VA is a leader in efforts to improve the quality of heart failure care. In a demonstration of its commitment to quality, the VA has provided data to the Centers for Medicare and Medicaid Services (CMS) so that CMS can report publically available comparisons of outcome and process measures for heart failure care to similar data from Medicare hospitals. CMS currently reports the following measures: 30-day all cause mortality; 30-day all cause readmission; measurement of left ventricular ejection fraction; use of an angiotensin converting enzyme inhibitor or angiotensin receptor blocker if the left ventricular ejection fraction is <40 percent; smoking cessation counseling; and patient education.[1]

Reducing 30-day Readmissions

During the last several years, heart failure performance measurement has focused on reducing 30-day readmissions, which occur in approximately 25 percent of patients according to the Hospital Compare website. Readmissions are a marker of many aspects of medical care, including severity of illness, quality of care by providers, quality of the health system in optimizing the transition of care, aggressiveness of care, patient adherence, and preference for location of care. The Department of Health and Human Services (DHHS) has made the assumption that $12 billion of the $15 billion dollars spent on readmissions could be saved due to eliminating preventable Medicare readmissions.

If the majority of readmissions were indeed preventable, it would be appropriate to hold hospitals accountable for these costs as is planned by CMS. However, preventability is extremely difficult to determine; in fact, researchers lack consensus on how to define a preventable readmission. Systematic reviews have shown wide variation in definitions of preventable admissions, with few studies of preventable readmissions deemed to be of high quality. For this reason, CMS has chosen 30-day all-cause readmission as the measure of quality. Recent studies suggest that the percent of Medicare readmissions that are preventable is much lower than the DHHS estimate of 75 percent. Researchers from Canada have estimated that the fraction of readmissions that is preventable is likely less than 20 percent.[2]

If better quality of care can improve the readmission rate, then randomized trials should be able to demonstrate this hypothesis. Indeed, more than a decade ago several studies found that the readmission rate for heart failure can be reduced through comprehensive discharge planning with a variety of interventions. Many of these interventions have since become standard care at VA hospitals. Whether more aggressive follow-up can continue to improve readmission rates is unknown. Paradoxically, researchers have found that readmission rates increase with closer follow-up as shown in a large VA randomized trial.

Searching for a Better Measure

Indirect evidence also suggests that 30-day readmission rates are now a poor measure of hospital or system quality of care. An analysis of Hospital Compare data showed that those hospitals with the highest readmission rates had improved mortality rates. The VA Chronic Heart Failure Quality Enhancement Research Initiative (QUERI) has shown that for Veterans over the last decade, as process of heart failure care measures improved, 30-day mortality rates also improved while readmission rates worsened slightly.[3]

If readmissions are a poor measure of quality, what would be better? First, the VA and DHHS could improve on the National Quality Forum endorsed measure for beta-blockers at discharge if the left ventricular ejection fraction is below 40 percent. Current VA use among candidates is about 70 percent. An emerging area for quality measurement is the safe use of aldosterone antagonists. These medications have reduced mortality and admissions for heart failure in randomized controlled trials, though improper monitoring can lead to dangerous hyperkalemia.

Perhaps more patient-centered outcome measures can be tested. From the patient's perspective, spending less time in the hospital is preferable, and if given the choice, many patients may prefer early discharge despite an increased risk of readmission. Thus, an alternative outcome measure to a 30-day readmission rate is total hospital days during the 30 days following the first day of admission. If resources were available for patient surveys, then a standardized measure of health status at 30 days following readmission would be even better. We should remember that our main goals with measurement of performance should be to improve patient length of life, quality of life, and efficient use of resources. Researchers should continually evaluate and revise performance measures as the health care system evolves.

References

1. Hospital Compare, Department of Health and Human Services, http://hospitalcompare.hhs.gov/.
2. Van Walraven, C. et al., "Incidence of Potentially Avoidable Urgent Readmissions and their Relation to All-cause Urgent Readmissions," *Canadian Medical Association Journal* 2011; 183(14).
3. Heidenreich, P.A., et al. "Divergent Trends in Survival and Readmission Following a Hospitalization for Heart Failure in the Veterans Affairs Health Care System 2002 to 2006," *Journal of the American College of Cardiology* 2010; 56(5):362-8.

Research Highlight

Improving Health Care for Veterans by Improving Health Care Quality Measures

Alex Sox-Harris, Ph.D., Center for Health Care Evaluation and Substance Use Disorder QUERI, VA Palo Alto Health Care System, Palo Alto, California

The health care that Veterans receive can be substantially improved by implementing validated health care quality measures because they define and motivate guideline-congruent care, reveal gaps in the continuum of care, and identify high- and low-performing facilities so that quality improvement efforts can be targeted. However, implementing quality measures without sufficient validation may promote poor or incomplete care, divert effort and attention from more important activities, and create skepticism and ill will toward the entire quality management enterprise.

Unfortunately, quality measures are often formulated and implemented without careful empirical validation or adequate appreciation for possible unintended consequences. Health services researchers can play a critical role by conducting validation studies of new and existing quality measures in order to refine their specifications and guide their interpretation and implementation. This article discusses three under-appreciated aspects of quality measure validation that, when addressed, can improve the quality of health care that Veterans receive.

The Ecological Fallacy is Under-Appreciated

Predictive validity refers to the association between antecedent quality indicators, particularly process quality measures, and subsequent quality indicators, particularly outcomes. Much of the research on the predictive validity of quality measures examines the associations between facility-level quality measure scores and average patient outcomes (e.g., mortality rates). Due to a phenomenon known as the ecological fallacy, such analyses tell you nothing about whether patients who receive quality-measure congruent care have better outcomes; this question can only be addressed with patient-level process and outcome data. In fact, the correlation between aggregated facility-level process data and average outcomes is often in the opposite direction of the patient-level association. This somewhat counter-intuitive fact has been long-appreciated in other scientific disciplines, but is only recently beginning to be recognized in the context of quality measure validation. Thus, in order to determine the predictive validity of a process quality measure—that is, if patients who satisfy the process criteria have better subsequent outcomes—it is necessary to use a model that contains patient-level process and outcome data.[1]

Quality Measure Specifications Need to be Checked

The ability to use available administrative data to accurately identify patients with particular characteristics and the occurrence of specific health care events is central to the validity of many quality measures. Specification validity refers to the sensitivity and specificity of the coding strategies used to identify and define the relevant patients and processes. In many areas of health care, available administrative data only approximately map onto consensus quality standards. For example, mental health care procedure codes such as "individual psychotherapy" do not specify the type or target of care (e.g., prolonged exposure therapy for PTSD). Although quality measure developers are often very creative in using combinations of diagnosis, procedure, and other codes to operationalize aspects of quality, the sensitivity and specificity of these coding strategies need to be verified through comparisons with other data sources such as chart review or direct observation. Studies of specification validity have revealed major problems in the specifications of long-established quality measures.[2,3] More studies of specification validity are needed to understand the limitations of the underlying coding strategies and improve them when possible.

More Research on Reactivity, Gaming, and Unintended Consequences is Needed

Quality measures vary in their vulnerability to improving *measured* performance without improving *actual* performance. For example, many quality measures have this form:

$$\frac{\text{Number of patients who receive a particular treatment}}{\text{Number of patients for whom the treatment is indicated}}$$

Facilities can improve measured performance by increasing the numerator, which is the intention, or by restricting the number of patients who qualify for the denominator. The validity of many quality measures relies on the invulnerability of the denominator to manipulation. This kind of "denominator management" can often be accomplished by examining if the overall proportion of all patients who meet the denominator criteria substantially changes once the quality measure is implemented, or if the proportion of patients qualifying for the denominator varies by facility to a surprising degree. For example, does the number of patients with a particular diagnosis, positive screen, or lab test change once a facility is held accountable to provide additional services to those patients? Studies such as this could determine if facilities are achieving high measured performance by restricting access to the denominator. Thus, creating denominator monitoring systems or quality measures that are less subject to manipulation are critically important to overall quality measure validity.

When these three under-appreciated aspects of health care quality measure validation are addressed, we can be more confident that the quality measures used in VA are having the intended effect: improving the quality of care that Veterans receive.

References

1. Finney, J.W. et al. "Why Health Care Process Performance Measures Can Have Different Relationships to Outcomes for Patients and Hospitals: Understanding the Ecological Fallacy," *American Journal of Public Health* 2011; 101(9):1635-42.
2. Harris, A.H. et al. "Validation of the Treatment Identification Strategy of the HEDIS Addiction Quality Measures: Concordance with Medical Record Review," *BMC Health Services Research* 2011; 11:73.
3. Harris, A.H. et al. "Are VHA Administrative Location Codes Valid Indicators of Specialty Substance Use Disorder Treatment?" *Journal of Rehabilitation Research and Development* 2010; 47(8):699-708.

Research Highlight

Are Financial Incentives to Health Care Providers Effective? Evidence from the Field

Laura Petersen, M.D., M.P.H., Houston Center for Quality of Care and Utilization Studies, Houston, Texas

Financial incentives have a powerful influence on the amount and type of health care provided to patients. Fee-for-service payments are associated with use of more (well-reimbursed) services; capitation payments are associated with fewer. Observations about the relationship between financing methods and use of services have influenced approaches to experiments with Accountable Care Organizations and other new models of care under the Affordable Care Act. Evidence regarding the impact of financial incentives on individual providers—as opposed to hospitals and delivery systems—is unclear.

Insights from the field of behavioral economics are beginning to inform approaches to a variety of human behavioral issues. For example, drivers don't necessarily respond to flashing signs and admonishments to slow down. The National Highway Traffic Safety Administration showed that drivers slowed dramatically when they were promised a $25 prize at the end of every week of safe driving, in combination with a penalty of six cents each time they went nine miles per hour above the speed limit. This surprisingly small penalty combined with the promise of a reward almost eliminated speeding. Audit and feedback comprised one interesting element of the intervention: every time drivers turned off the ignition, they were informed about the status of their incentive payment.

Lessons from Behavioral Economics

Several of us at the Houston HSR&D Center of Excellence are experimenting with applications of behavioral economics to primary care. VA has a robust quality measurement and monitoring program and a strong culture of reporting and accountability, including a physician bonus program. Bearing in mind behavioral economics lessons from other sectors such as traffic safety, are further improvements in performance possible?

With funding from VA HSR&D, we assessed the effectiveness of financial incentives to overcome clinical inertia and improve hypertension guideline adherence. We enrolled primary care physicians from 12 VA primary care clinics in 5 networks. We cluster-randomized physicians at the clinic level to one of four arms: incentives to individual physicians; incentives to health care teams or "groups"; incentives to both; or control (no incentives). Groups consisted of physicians and non-physician personnel (e.g. nurses, pharmacists, and clerks). Participants in all four arms received education and audit and feedback on their performance. Of course, there are legitimate concerns about unintended consequences of applying individual incentives to health care; following the speed limit is quite different from the practice of medicine. Therefore, our work incorporates a careful mixed methods assessment of unintended consequences.

We calculated the proportion of sampled patients meeting each of two measures: 1) receiving guideline-recommended antihypertensive medications; and 2) achieving the guideline-recommended blood pressure threshold OR receiving an appropriate response to uncontrolled blood pressure. These two measures were rewarded independently; the maximum incentive was paid for achieving both. To dampen incentives for adverse selection of patients with resistant hypertension, we rewarded appropriate actions in response to elevated blood pressure, regardless of whether the target blood pressure was achieved.[1] The five participating VA Network partners provided the incentive payment pool. The maximum total individual performance reward was $3,640.

Participants could view their customized audit and feedback reports, including individual performance, earnings, and future performance goals, on a password-protected website. Intervention arm participants were more likely than controls to view their feedback (68 percent v. 25 percent, respectively, p=0.001) and viewed it 4.5 times more often (p=0.001). In a multivariate model accounting for ceiling effects, a study physician with a panel of 1,000 hypertensive patients would achieve the performance measure of controlled blood pressure or appropriate clinical response to uncontrolled blood pressure on 61 more patients after one year of exposure to the individual incentive compared to controls (p=0.003).[2] We now are examining whether incentivizing hypertension care produced neglect of other aspects of health care (also referred to as the "spotlight effect").

Impact on Guideline Adherence

Why might financial incentives work to improve guideline adherence? Financial incentives for individual effort and task performance might amplify the positive effects of educational interventions and goal-setting performance feedback reports. According to Bandura's self-efficacy theory, incentives work by piquing an individual's interest in a task, leading to greater effort at performing the task and ultimately to an increased sense of self-efficacy.[3] This conceptual approach fits well with our design of rewards for following a guideline (a fairly discrete set of tasks), as opposed to complex problem solving (e.g. diagnosing the etiology of abdominal pain). The goal of the incentive, rather than to coerce, is to ignite motivation. Our qualitative and quantitative

Continued on page 8

FORUM
Translating research into quality health care for Veterans

Center for Information Dissemination and Education Resources
VA Boston Healthcare System (152C)
150 South Huntington Avenue, Boston, MA 02130-4893

Financial Incentives – continued from page 7

data suggest that the incentive increased participant interest in viewing their feedback and in improving their performance. Of course, there are myriad reasons, including professionalism and intrinsic motivation, for physicians to do a good job. But our findings suggest that even in a system such as VA—with its strong culture of measurement, reporting, and accountability—financial incentives that signal health care providers about the importance of a discrete task have the potential to improve performance.

References

1. Petersen, L.A. et al. "Design, Rationale, and Baseline Characteristics of a Cluster Randomized Controlled Trial of Pay for Performance for Hypertension Treatment: Study Protocol," *Implementation Science* 2011; 6(1):114.
2. Petersen, L.A. et al. "Do Financial Incentives for Guideline Adherence Improve Care of Hypertension in the VA Primary Care Setting? A Multi-Site Randomized Trial" Presented; HSR&D Annual Meeting, 2012.
3. Bandura, A. 1997. *Self-efficacy: the Exercise of Control;* W.H. Freeman and Company: New York, NY.