Pitfalls of Converting Practice Guidelines Into Quality Measures
Lessons Learned From a VA Performance Measure

Louise C. Walter, MD
Natalie P. Davidowitz, BA
Paul A. Heineken, MD
Kenneth E. Covinsky, MD, MPH

There is great demand for accurate measures of quality to determine whether consistent high-quality medical care is being provided across health care systems. The US Department of Veterans Affairs (VA) manages the largest health care system in the United States and is considered a leader in quality measurement.1 To assess quality of care at each of its medical centers, the VA calculates scores on a broad range of clinical performance measures based on standardized chart review by external auditors.2 These measures are generally based on adherence rates to guideline-based performance measures. A recent study concluded that higher adherence rates to guideline-based performance measures were evidence of improved quality of care across the VA health care system.3 At the local VA level, there is considerable pressure to score well on these measures, which are widely publicized.3,4 This pressure was recently illustrated at the San Francisco VA Medical Center (SFVAMC) when clinicians were told that the medical center’s rate of colorectal cancer screening in 2002 of 58% failed to meet the national VA target rate of 65%. Clinicians were strongly encouraged to screen more patients and were told that failure to increase screening rates could result in financial penalties for the medical center. Intrigued with the interpretation that lower screening rates indicated a quality-of-care problem, we critically assessed the colorectal cancer screening performance measure by re-

The Department of Veterans Affairs (VA) manages the largest health care system in the United States, and the Institute of Medicine has recommended that many practices of VA quality measurement be applied to the US health care system as a whole. The VA measures quality of care at all of its sites by assessing adherence rates to performance measures, which generally are derived from evidence-based practice guidelines. Higher adherence rates are used as evidence of better quality of care. However, there are problems with converting practice guidelines, intended to offer guidance to clinicians, into performance measures that are meant to identify poor-quality care. We suggest a more balanced perspective on the use of performance measures to define quality by delineating conceptual problems with the conversion of practice guidelines into quality measures. Focusing on colorectal cancer screening, we use a case study at 1 VA facility to illustrate pitfalls that can cause adherence rates to guideline-based performance measures to be poor indicators of the quality of cancer screening. Pitfalls identified included (1) not properly considering illness severity of the sample population audited for adherence to screening, (2) not distinguishing screening from diagnostic procedures when setting achievable target screening rates, and (3) not accounting for patient preferences or clinician judgment when scoring performance measures. For many patients with severe comorbid illnesses or strong preferences against screening, the risks of colorectal cancer screening outweigh the benefits, and the decision to not screen may reflect good quality of care. Performance measures require more thoughtful specification and interpretation to avoid defining high testing rates as good quality of care regardless of who received the test, why it was performed, or whether the patient wanted it.
viewing the medical records of all patients previously audited for colorectal cancer screening at SFVAMC in 2002. Our findings illustrate conceptual problems with the conversion of practice guidelines into quality measures, which have potentially important implications for quality measurement in the VA and other health care systems.

The national VA performance measure for colorectal cancer screening was developed from practice guidelines based on trials demonstrating decreased colorectal cancer mortality in patients offered screening. The adoption of practice guidelines based on strong scientific evidence lends clinical credibility that higher screening rates could reflect improved quality of care in the VA system. However, performance measures are not the same as practice guidelines (Table). Practice guidelines are intended to offer guidance by providing information to clinicians to help them decide how best to care for patients, allowing for clinical judgment and patient preferences. In contrast, performance measures are quantitative tools, such as rates, used to set standards that, if not met, almost certainly identify poor-quality care.9

As we conducted our in-depth review of the VA’s 2002 audit of colorectal cancer screening at SFVAMC, we identified several pitfalls that can occur when converting practice guidelines into performance measures. These include problems selecting the appropriate target population, determining target screening rates, and measuring screening performance. Such problems seem particularly important to understand given the recent recommendation of the Institute of Medicine to apply many VA performance measures to the US health care system as a whole.

Target Population: Does Severity of Illness Matter?

Multiple evidence-based practice guidelines strongly recommend screening persons aged 50 years or older for colorectal cancer with fecal occult blood testing annually, sigmoidoscopy every 5 years, or colonoscopy every 10 years. The VA translated this guideline into a national performance measure for all VA medical centers that computes the percentage of eligible patients receiving timely colorectal cancer screening by any of these methods. The target population included all patients aged 52 years or older who were seen at least once during the fiscal year of the audit in 1 of 8 outpatient medical clinics. No upper age limit was designated. Patients were excluded from the eligible target population only if they (1) had a documented diagnosis of cancer of the esophagus, liver, or pancreas; (2) were enrolled in a hospice program; or (3) had a documented life expectancy of less than 6 months. Documentation by the clinician that colorectal cancer screening was not medically indicated for other reasons was not considered an exclusion criterion by the auditors.

A sample of 229 patient charts was selected by auditors during fiscal year 2002 to calculate the SFVAMC’s rate of colorectal cancer screening. However, this sample was not a simple random sample of those meeting inclusion criteria for the colorectal cancer screening performance measure. Since the VA was also auditing adherence to performance measures for chronic diseases, such as chronic obstructive pulmonary disease and ischemic heart disease, the VA decided that as a national policy it would be more efficient and cost-effective to audit patient charts in which multiple performance measures could be evaluated simultaneously. Therefore, auditors searched national VA inpatient and outpatient databases for specific International Classification of Diseases, Ninth Revision (ICD-9) codes to oversample charts of patients with specific comorbid illnesses.

We found that the minimal exclusion criteria and sampling strategy that oversamples chronically ill patients resulted in a skewed target population that was elderly with multiple comorbid illnesses. For example, of the 229 patient charts audited at the SFVAMC during fiscal year 2002 for performance of colorectal cancer screening, 35% were aged 75 years or older and 24% had a severe comorbid illness (Figure 1). Some examples of patients in this sample included an 89-year-old woman with severe Alzheimer disease–related dementia (Mini-Mental State Examination score, 7/30), a 94-year-old man with metastatic prostate cancer, and a 76-year-old man with end-stage renal disease who was receiving dialysis, who had had a recent leg amputation, and who died 2 months prior to the audit date. Sixty-seven percent of all patients had a Charlson comorbidity-age combined risk score of 4 or more, which is associated with greater than 50% 10-year mortality.

Since cancer screening offers potential benefits only to patients who are likely to live long enough to develop symptoms from an asymptomatic screen-detectable cancer, patients with severe comorbid illnesses associated with a life expectancy of less than 5 to 10 years are
unlikely to benefit from screening.\textsuperscript{15} Furthermore, the risk of complications from colorectal cancer screening increases with increasing comorbidity.\textsuperscript{15,16} As a result, the sample that the VA selected in which to measure performance of colorectal cancer screening is not the population in which failure to screen almost certainly identifies poor-quality care. In fact, for a substantial proportion of patients chosen for the audit, the risks of screening would outweigh the potential benefits.\textsuperscript{15}

Severity of illness is rarely considered when comparing rates of processes of care such as cancer screening.\textsuperscript{17,18} It is generally assumed that process-based measures are targeted to patients in whom an intervention has been shown to be efficacious or there are strong reasons to extrapolate evidence of efficacy. However, the VA did not properly consider illness severity or prognosis when implementing its colorectal cancer screening performance measure, which was targeted to patients who were much sicker than the healthy volunteers who participated in randomized trials of screening.\textsuperscript{5,7}

**Target Screening Rates: Are Higher Rates Always Better?**

Performance measures often attempt to address that most interventions are not clinically appropriate for all patients in a target population by setting performance goals at less than 100%. This provides an allowance for exceptions without the expense and labor required to audit charts for reasons when an intervention is not indicated. However, the target rates that are ultimately set are not derived from evidence-based processes or practice guidelines. At the VA, the Office of Quality and Performance sets target screening rates by (1) evaluating goals and results from other systems, such as Healthy People 2010\textsuperscript{19} and HEDIS (Health Plan Employer Data and Information Set),\textsuperscript{20} and (2) using a continuous quality improvement model to reset target goals each year to the screening rate of the top 20% of VA networks from the previous year.\textsuperscript{21} This is based on the principle of benchmarking, in which a health care system tries to model the performance of its top-performing facilities by making it the standard with which others are compared.\textsuperscript{22} Based on these methods, the VA set the “successful” goal for colorectal cancer screening at 65% for fiscal year 2002 and the “exceptional” goal at 71%.\textsuperscript{13}

However, there are several problems with this approach to setting target rates. First, the tests used to screen for colorectal cancer are also frequently used as diagnostic tests to work up anemia, hematochezia, etc. Since auditors do not consider the indication for the test, the target “screening” rate may be driven by high rates of diagnostic testing at some VA medical centers. For example, while auditors found that SFVAMC had a colorectal cancer screening rate of 58% (132/229) in fiscal year 2002, we found that 42% (55/132) of the “screened” patients were tested for diagnostic rather than screening purposes (FIGURE 2). Therefore, despite the high illness severity of the audited patients at the SFVAMC, the “screening” rate was as high as it was because these patients often had gastrointestinal problems requiring colonoscopy in the prior 10 years. While the percentage of diagnostic tests would likely decrease if the audited population was younger and healthier, the methods for setting target rates should consider test indication since tests performed for diagnostic purposes are not informative about the quality of screening.

Second, even if tests performed only for the purpose of screening were used to set target screening rates, the process of resetting rates according to the top 20% of VA networks carries the implicit judgment that higher screening rates reflect better-quality care. However, a more clinically sensible goal would be to optimally target screening to those most likely to benefit rather than achieve as high a screening rate as possible, since cancer screening should vary according to prognosis and patient preferences.\textsuperscript{23} In the extreme, if high target goals push clinicians to screen patients with limited life expectancies, performance measures can lead to unnecessary procedures and distract from more important aspects of care.\textsuperscript{24,25}

**Measuring Performance: Should Patient Preferences and Clinical Judgment Be Considered?**

Process-based performance measures are usually scored as a simple percentage in which the number of persons who receive an intervention (numerator) is divided by the number of eligible patients (denominator). Patient preferences and clinical judgment are sometimes addressed by the way these measures are scored when a patient declines an intervention or when a clinician documents that an intervention is “not medically indicated.” However, some performance measures ignore these issues despite practice guidelines that recommend that screening decisions consider patient preferences and severity of comorbid disease.\textsuperscript{11,12}

For example, patients who decline or do not adhere to screening, as well as patients in whom their clinician documents that screening is not medically indicated, are viewed by the national VA performance measure as eligible patients who failed to get screened; ie, they are counted in the denominator but not...
the numerator. There are several justifications for this approach. First, while patient preferences against screening “count against” a medical center, this is partially taken into account by setting target goals at less than 100%..

Second, from a system perspective, performance measures are intended to give an accurate picture of the relative quality of care delivered by a facility, even if they are not accurate for some individuals.

However, when a performance measure does not validate responsible clinical decision making, it may undermine clinicians’ confidence and trust in performance measurement. It also minimizes the importance of tailoring medical recommendations to the needs and preferences of individual patients, an essential component of patient-centered care. These concerns are especially acute for colorectal cancer screening, a procedure that depends heavily on patient preferences, adherence, and severity of illness.

Our experience at the SFVAMC illustrates the importance of these issues. Of the 81 patients who did not receive colorectal cancer screening, 38 (47%) had documentation that they declined screening, 10 (12%) failed to complete or show up for screening tests, and 9 (11%) had notes detailing why screening was not medically indicated (Figure 2). While this clinical detail cannot be captured when performance measures are based on ICD-9 codes from administrative data, VA auditors already review medical charts that contain this information. Sometimes medical charts may lack details about why patients refuse or do not adhere to screening, making it difficult to determine when a patient’s decision against screening reflects problems with patient-clinician communication or the patient’s informed preference. However, while a patient’s decision against screening may not always equate with good-quality care, it is currently labeled unequivocally as poor-quality care.

Another, more obvious measurement error can occur when auditors simply miss the documentation of screening in the medical records. Since documentation of screening can occur in multiple places in the electronic medical record and documentation can occur as far back as 10 years ago for a colonoscopy, it is not surprising that auditors missed some documented screening (Figure 2). Two authors (L.C.W. and N.P.D.) spent an average of 20 minutes per chart searching for documentation. Extensively reviewing charts is time-consuming and suggests that problems with accuracy may increase when the time available to search for documentation progressively decreases as the list of performance measures grows.

**Recommendations**

Despite the pitfalls that may occur in converting a practice guideline into a performance measure, the potential benefits of performance measures derived from evidence-based guidelines should not be ignored. The quality of medical care can be improved in many areas, and performance measures are tools that can help achieve selected goals.

Certainly, the national VA colorectal cancer screening performance measure spotlights the importance of colorectal cancer screening, which is often underused in patients who may benefit. However, pitfalls of converting practice guidelines into performance measures must be recognized and avoided. If one is not careful, performance measures can become scientifically insupportable algorithms that classify high rates of a procedure as good care regardless of who received it, why it was performed, or whether the patient wanted it. In addition, other authors have suggested that such measures may have harmful unintended consequences to patients (adverse outcomes from unnecessary procedures and diversion from beneficial uses of time with their clinician), clinicians (decreased morale and professionalism), and organizations (consumption of scarce resources).

Since performance measures are meant to distinguish between good- and poor-quality care rather than to be advisory, they need to be applied to a more narrowly defined population than practice guidelines. The target population for a performance measure should include only those for whom good evidence exists that the benefits significantly outweigh the harms and those who do not refuse the intervention. These specifications are especially critical when defining the quality of cancer screening, because a patient’s health status and pref-
ferences appropriately influence the receipt of cancer screening more than they influence some treatment interventions. For example, health status and preferences are more likely to influence receipt of colorectal cancer screening than receipt of a β-blocker after a myocardial infarction. Therefore, rather than measuring the percentage of patients screened for cancer during a recommended interval, a better measure of the quality of screening would be to measure thoughtful decision making that includes prognosis and patient preferences. Veterans Affairs auditors already review medical charts and could easily abstract this information if clinicians documented their discussions and recommendations in a standardized way.

If there is concern that clinicians would inappropriately document a screening test as “not medically indicated,” a few randomly chosen charts could be reviewed in depth to ensure that the reasoning behind a screening decision was clinically sound. In fact, in-depth chart reviews and even patient interviews should regularly support the quality of care, since sim- ply reporting screening rates conceals measurement forward. The difficulties measuring the quality of colorectal cancer screening demonstrate that even well-intentioned performance measures can have pitfalls and emphasize that performance measurement is a young technology that requires further evaluation before it can achieve its full role in improving patient care.

Author Contributions: Study concept and design: Walter, Heineken, Covinsky. Acquisition of data: Walter, Davidowitz. Analysis and interpretation of data: Walter. Drafting of the manuscript: Walter, Davidowitz, Heineken, Covinsky. Critical revision of the manuscript for important intellectual content: Walter, Heineken, Covinsky. Obtained funding: Walter. Administrative, technical, or material support: Davidowitz, Heineken.

Funding/Support: Dr Walter is a recipient of the Veterans Affairs Career Development Award in Health Services Research and Development. Dr Covinsky was supported in part by independent investigator award KOHS00006-01 from the Agency for Healthcare Research and Quality and is a Paul Beeson Faculty Scholar in Aging Research.

Role of the Sponsors: The funding sources had no role in the design, conduct, analyses, or decision to publish this study.

REFERENCES