Design and Use of Performance Measures to Decrease Low-Value Services and Achieve Cost-Conscious Care

David W. Baker, MD, MPH; Amir Qaseem, MD, PhD, MHA; P. Preston Reynolds, MD, PhD; Lea Anne Gardner, PhD, RN; and Eric C. Schneider, MD, MSc, on behalf of the American College of Physicians Performance Measurement Committee*

Improving quality of care while decreasing the cost of health care is a national priority. Several organizations that have traditionally focused on increasing use of beneficial services have intensified their efforts to decrease the use of low-value health care services. In 2006, the National Committee on Quality Assurance proposed a quality performance criterion for overuse of spine imaging (1). In 2008, the National Priorities Partnership identified “overuse” as 1 of 6 national health care priorities (2). More recently, the American College of Physicians launched its High-Value Care Initiative (3), which seeks to help physicians and patients understand the benefits, harms, and costs of interventions and whether services provide good value (4–6). For example, the American College of Physicians’ paper on high-value care for low back pain advocates using diagnostic imaging only when patients have progressive neurologic deficits or signs or symptoms suggestive of a serious or specific underlying condition; routine imaging is otherwise considered to be low-value (4).

Just as we need performance measures to assess underuse of high-value services, we need valid, evidence-based measures of overuse. For example, at the same time that we should be measuring the proportion of patients aged 50 to 75 years who have been screened for colorectal cancer, we should be assessing the proportion of patients older than 75 years who had colorectal cancer screening that was not indicated. Performance measures for low-value services have the potential to be an important lever for changing clinician behavior through feedback, public reporting, clinical decision support, and financial incentives. This paper gives an overview of performance measures for low-value services, provides specific examples of possible measures to assess use of low-value services, and discusses how these measures can be used in clinical practice and policy.

LOW-VALUE SERVICES

This discussion includes two categories of interventions: 1) those for which the harms likely exceed the benefits and 2) those that may provide benefits, but for which a quantitative assessment of their benefits and costs by a multistakeholder group (patients, clinicians, and policymakers) suggests that the tradeoff between health benefits and expenditures is undesirable. Use of services for which the harms likely exceed the benefits has been defined by the Institute of Medicine as overuse. The RAND “appropriateness methodology” defines a test or treatment as “inappropriate” if the risk exceeds the benefit of the procedure for a specific indication. An example of such a service is colorectal cancer screening for patients older than 85 years; any small benefit in detecting polyps or early colorectal cancer is outweighed by the predictable and unavoidable possibility of harm from colonic perforation during the procedure and the competing risk for death from other causes (7).

The second category includes services for which the risk-to-benefit ratio is uncertain, as well as those that have a definable benefit but the benefit is judged to be outweighed by the relative harms and cost of the services (for example, by a cost-effectiveness analysis). There is no universally accepted methodology or bright line that defines the point at which a service has so little value that it should...
not be done (8); this will ultimately be a societal decision that depends on how much money we are willing to spend on health care, along with societal priorities. There are many examples of services with little or no value, such as screening for cervical cancer in low-risk women aged 65 years or older and in women who have had a total hysterectomy (uterus and cervix) for benign disease, performing imaging studies (rather than a high-sensitivity D-dimer measurement) as the initial diagnostic test in patients with low pretest probability of venous thromboembolism, and screening for chronic obstructive pulmonary disease with spirometry in individuals without respiratory symptoms (6).

**Types of Measures and Measurement Approaches**

**Direct Measures**

A direct measure makes a judgment about whether an intervention was of low-value on the basis of the unique clinical circumstances of each eligible patient. For example, a direct measure of imaging for patients with acute low back pain would determine whether the patient had an imaging test that is typically of low-value and whether unusual circumstances justified the imaging test (such as a history of cancer). A clinician’s performance is measured as the proportion of all eligible patients for whom he or she is responsible who received the low-value service and did not have extenuating circumstances. The theoretical optimal performance is 100% of patients not getting the test or intervention or having a documented justification for why they should get the service. However, direct measures require access to detailed clinical information to make these judgments, and these data are often not easily obtained.

**Indirect Measures**

An indirect measure evaluates use rates, and exceptionally high use rates are assumed to indicate that a provider (or provider group) frequently uses services of low-value. Indirect measures must be used when specific clinical criteria have not been defined to directly measure use of low-value services or when data sources containing the highly detailed clinical information required for direct measurement are not available. For example, administrative data and electronic health record data can be used to measure rates of diagnostic imaging for specific conditions. The ideal use rate of diagnostic imaging is not known because the data sources may be unable to identify all patients for whom an imaging study is actually justified and appropriate (such as patients with back pain who are at high risk for cancer or spinal infection). Indirect measures therefore use a normative approach and compare clinicians’ use rates for a service to their peers. Because use data are ubiquitous and rates are relatively easy to calculate, indirect measures can assess many low-value services.

Many studies have supported the validity of relying on use rates as indirect measures of use of low-value services; these studies have shown wide variations in use of health care without apparent improvements in health outcomes (9–11). However, the assumption that very high rates represent more frequent use of low-value services may not always be true. If the number of people who truly need services varies substantially across clinicians, then raw rates of service use may not always be valid proxies for rates of unnecessary use. Thus, some caution is necessary. One study found that variations in use of coronary angiography, carotid endarterectomy, and upper gastrointestinal tract endoscopy across geographic areas were weakly associated with or not associated with rates of inappropriate use (12).

Interpretation of indirect measures is even more challenging when the proper use of a diagnostic test depends on the a priori likelihood of the disease being considered, and that probability may range from near 0 to almost 100%. For example, when is it appropriate to perform computed tomography to assess pulmonary emboli in a patient presenting with chest pain? At what a priori probability does the risk from radiation exposure exceed the likely benefit? Thus, even when normative data are available, use rates are often difficult to interpret in these circumstances. The Centers for Medicare & Medicaid Services Hospital Compare Web site reports hospitals’ rates of follow-up mammography or ultrasonography within 45 days after screening mammography (13). To help patients interpret hospitals’ rates, the site says, “A number that is much lower than 8% may mean there’s not enough follow-up. A number much higher than 14% may mean there’s too much unnecessary follow-up” (italics added). Whether outliers on this type of measures of use are truly overusing or underusing services is not clear and requires further study.

**Using Rates of Negative Results to Improve Indirect Measures of Use of Low-Value Services**

One potential way of improving gross use rates as indirect measures of low-value service use is to examine the rates at which results of diagnostic tests are determined to be negative (that is, no abnormality is found related to the presenting symptom). If a diagnostic test is used too often for low-risk patients, this will result in 1) a higher than expected rate of use and 2) a higher than normal rate of negative test results. For example, only one third of patients without known coronary artery disease were found to have obstructive lesions when they underwent elective cardiac catheterization (14). Similarly, a study of 28,177 patients who had revascularization found that 61% of patients with percutaneous coronary intervention and 51% of patients with coronary artery bypass grafting had testing for ischemia (most often nuclear imaging) by 24 months (15). However, of patients tested, only 11% had subsequent cardiac catheterization and only 5% had repeated revascularization (15). Thus, the a priori probability of obstruction was very low, suggesting that most persons tested had weak or no indications.
There are many other situations for which the rate of negative test results may be more helpful than gross utilization rates. For example, the Centers for Medicare & Medicaid Services Hospital Compare Quality Measures report rates of follow-up imaging after mammography. Some women have equivocal findings on mammography and require additional imaging to determine whether biopsy is needed. The proportion of women who need additional imaging after screening mammography depends on the case mix, especially if the center is a referral center that may be performing screening mammography in women with a history of breast cancer who are at higher risk for new lesions. However, the rate of additional imaging also depends on how risk-averse the interpreting radiologist is. A high rate of additional imaging coupled with a low rate of abnormal test results and subsequent biopsy would provide additional evidence of overuse.

Although using rates of negative test results as indirect measures of overuse has distinct advantages compared with using only crude rates of test ordering, changes in reporting standards would be needed for this to be possible. Administrative claims data lack the test results necessary to determine the rate of negative results. Even with electronic health records, test results are often stored as text rather than in discrete fields that could be queried. Nevertheless, this method holds promise for improving the validity and interpretability of using use rates as a way of measuring use of low-value services and should be prioritized for further evaluation.

**The Evidence Base for Creating Performance Measures for Low-Value Services**

Ideally, performance measures should be based on rigorous study designs (for example, randomized, controlled trials) that assessed the benefits, risks, and costs of interventions. However, to develop performance measures for low-value services, we will probably need to use data from different types of research design and methods, including subgroup analyses from clinical trials, cohort studies, cost-benefit analyses, and cost-effectiveness analyses. For example, a study reported that the net clinical benefit of anticoagulation among patients with atrial fibrillation and CHADS2 (congestive heart failure, hypertension, age >75 years, diabetes mellitus, and prior stroke) scores of 0 or 1 were “essentially zero” (16). This could be used to create a measure of anticoagulation use in this subgroup for whom anticoagulation has little or no value.

**Individual Versus Group-Level Performance Measurement**

Performance measures for use of low-value services will probably need to be applied at the group level, such as a hospital or multispecialty group. Many individual clinicians may not see enough patients with the target conditions within the measurement interval to allow reliable measurement of differences in use. In addition, primary care physicians and specialists are often involved in decision making, and both should be held accountable rather than just the person who ordered the test. For example, some specialists may request that an imaging test be completed before they will see a patient. A gastroenterologist may recommend repeated colonoscopy for colonic polyp surveillance at a shorter interval than suggested by national guidelines, and a primary care physician may then feel obligated to follow that recommendation. Small-area variation studies have shown that regional and probably organizational cultures affect health service use.

However, applying these measures at the group level may be problematic if we consider the large number of physicians who are in solo or small-group practices. These physicians must refer into a hospital or imaging center for diagnostic testing. To combine their use rates with those of other unaffiliated physicians may be unfair, and the hospital itself may believe it is unfair to be held responsible for the use patterns of referring physicians. Nevertheless, group-level measures have the advantage of creating communities of clinicians with shared responsibility for decreasing use of low-value services.

**Applying Performance Measures to Improve Value**

Just as with other performance measures, those for low-value services can be used in a variety of ways to improve quality and health care value. A commonly used quality improvement strategy is audit and feedback, in which performance is measured and summaries of performance are given to clinicians (17). Audit and feedback seem to modestly improve quality of care, especially when performance is mediocre or poor (17). Less is known about the value of audit and feedback for measures of use of low-value services (18–20).

A second possible use is public reporting. The Centers for Medicare & Medicaid Services has continued to expand the number and type of public performance reports available over the Internet. This currently includes tools to compare hospitals, nursing homes, and dialysis centers, and plans are under way for providing information and tools to compare individual physicians (13, 21). However, little is known about whether public reporting of performance measures for use of low-value services will change use rates; such changes could occur if clinicians alter their practice patterns or if patients choose clinicians who seem to order services more judiciously (22). Previous studies suggest that patients are not familiar with public reports on quality (22, 23). More intensive dissemination efforts are needed if public reporting of overuse measures is to be effective. These will need to be coupled with patient education about the lack of need for specific services (such as imaging for low back pain and cervical cancer screening after a...
hysterectomy) and communicated through multiple media sources over a sustained period.

Financial incentives to discourage use of low-value services are also likely to be used (24). If direct measurement is possible, payors could deny payments for interventions for which patients do not meet specific criteria. For example, Medicare has established specific circumstances under which they will cover continuous positive airway pressure machines for patients with obstructive sleep apnea (25). Payors could also require high copayments from patients. When indirect measures are used (that is, when clinical indicators are not available or the data required are not accessible), payors could increase payments for clinicians with low rates of use of low-value services or decrease payments for clinicians with high rates. This is similar to other pay-for-performance programs. However, pay-for-performance programs that use indirect measures indiscriminately to address use of low-value services risk decreasing use so that some patients who need services do not receive them. For example, even though rates of coronary artery bypass grafting are lower in the United Kingdom, a substantial proportion of procedures are still judged to be inappropriate (26, 27). This raises concerns that efforts to decrease crude use rates without simultaneous efforts to increase appropriate use could be harmful.

Electronic health records with advanced clinical decision support (CDS) may also be used to decrease use of low-value services. The logic for direct measures of use of low-value services can be programmed into electronic health records to create point-of-care alerts and CDS tools to help clinicians decide whether they are about to order a low-value intervention. Electronic CDS reminders have been shown to improve performance for many, but not all, targets (28–31). Far fewer studies have examined whether CDS reminders and other tools can decrease unnecessary use. Most studies have addressed overprescribing antibiotics for upper respiratory tract infections (20). Two recent studies suggest that order-entry CDS tools can decrease inappropriate ordering of radiology tests and decrease the rate of growth of outpatient computed tomography (32, 33).

**Conclusion**

The first step in addressing the high cost of health care should be decreasing use of interventions that provide little or no benefit and are of low value. During the past 2 decades, we have learned a great deal about how to rigorously develop quality measures and how to use them to improve care. This knowledge can be applied to develop performance measures for low-value care. Measurements can be made directly, to evaluate care for individual patients, or indirectly, by analyzing use data at an aggregate level. Each measurement approach has strengths as well as limitations. Performance measures for low-value care need to be developed and tested with the same rigorous methods as performance measures for underuse of services; however, the evidence base used to develop measures will differ substantially. Evidence-based performance measures for low-value services can help motivate physicians to provide high-value care to their patients.

From the Feinberg School of Medicine Chicago, Illinois; American College of Physicians, Philadelphia, Pennsylvania; University of Virginia School of Medicine, Charlottesville, Virginia; ECR Institute, Plymouth Meeting, Pennsylvania; and RAND, Boston, Massachusetts.

**Financial Support:** Financial support for the development of this paper comes exclusively from the American College of Physicians operating budget.

**Potential Conflicts of Interest:** Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M12-0480.

**Requests for Single Reprints:** Amir Qaseem, MD, PhD, MHA, American College of Physicians, 190 N. Independence Mall West, Philadelphia, PA 19106; e-mail, aqaseem@acponline.org.

Current author addresses and author contributions are available at www.annals.org.

**References**

1. National Committee for Quality Assurance. HEDIS. Technical Specifications. 2006. Accessed at www.ncqa.org/tabid/59/default.aspx on 9 April 2012.

2. National Priorities Partnership. National Priorities and Goals: Aligning Our Efforts to Transform America’s Healthcare. Washington, DC: National Quality Forum; 2008.

3. American College of Physicians. High value care. 2010. Accessed at www.acponline.org/practice_information/resources/highvalue.htm on 9 April 2012.

4. Chou R, Qaseem A, Owens DK, Shekelle P; Clinical Guidelines Committee of the American College of Physicians. Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. Ann Intern Med. 2011;154:181-9. [PMID: 21282698]

5. Owens DK, Qaseem A, Chou R, Shekelle P; Clinical Guidelines Committee of the American College of Physicians. High-value, cost-conscious health care: concepts for clinicians to evaluate the benefits, harms, and costs of medical interventions. Ann Intern Med. 2011;154:174-80. [PMID: 21282697]

6. Qaseem A, Alguire P, Dallas P, Feinberg LE, Fitzgerald FT, Horwich C, et al. Appropriate use of screening and diagnostic tests to foster high-value, cost-conscious care. Ann Intern Med. 2012;156:147-9. [PMID: 22250146]

7. U.S. Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2008;149:627-37. [PMID: 18838716]

8. American College of Physicians. How can our nation conserve and distribute health care resources effectively and efficiently? 2011. Accessed at www.acponline.org/advocacy/where_we_stand/policy/health_care_resources.pdf on 9 April 2012.

9. Fisher ES, Wennberg JE, Stukel TA, Skinner JS, Sharp SM, Freeman JL, et al. Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. Health Serv Res. 2000;34:1351-62. [PMID: 10654835]

10. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care. Ann Intern Med. 2003;138:288-98. [PMID: 12585826]

11. Wennberg JE, Freeman JL, Shelton RM, Bubolz TA. Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. N Engl J Med. 1989;321:1168-73. [PMID: 2677726]

12. Chassin MR, Kosecoff J, Park RE, Winslow CM, Kahn KL, Merrick NJ, et al. Does inappropriate use explain geographic variations in the use of health care?
Performance Measures to Decrease Low-Value Services

13. Centers for Medicare & Medicaid Services. Hospital Compare Web site. Accessed at www.hospitalcompare.hhs.gov on 9 April 2012.

14. Patel MR, Peterson ED, Dai D, Brennan JM, Redberg RF, Anderson HV, et al. Low diagnostic yield of elective coronary angiography. N Engl J Med. 2010;362:886-95. [PMID: 20220183]

15. Shah BR, Cowper PA, O’Brien SM, Jensen N, Drawz M, Patel MR, et al. Patterns of cardiac stress testing after revascularization in community practice. J Am Coll Cardiol. 2010;56:1328-34. [PMID: 20888523]

16. Singh D, Chang Y, Fang MC, Borowsky LH, Pomerancky N, Udaltsova N, et al. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. Ann Intern Med. 2009;151:297-305. [PMID: 19721017]

17. Ivers N, Jamtvedt G, Florøtort S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012:6:CD000259. [PMID: 22696318]

18. French SD, Green S, Buchbinder R, Barnes H. Interventions for improving the appropriate use of imaging in people with musculoskeletal conditions. Cochrane Database Syst Rev. 2010:CD006094. [PMID: 20091583]

19. Agency for Healthcare Research and Quality. Closing the quality gap: a critical analysis of quality improvement strategies Volume 4—antibiotic prescribing behavior. Structured abstract. January 2006. Accessed at www.ahrq.gov/clinic/pf/medigaptp.htm on 9 April 2012.

20. Steinman MA, Ranji SR, Shojania KG, Gonzales R. Improving antibiotic selection: a systematic review and quantitative analysis of quality improvement strategies. Med Care. 2006;44:617-28. [PMID: 16799356]

21. Centers for Medicare & Medicaid Services. Physician Compare. Accessed at www.medicare.gov/find-a-doctor/provider-search.aspx on 9 April 2012.

22. Schneider EC, Lieberman T. Publicly disclosed information about the quality of health care: response of the US public. Qual Health Care. 2001;10:96-103. [PMID: 11389318]

23. Schneider EC, Epstein AM. Use of public performance reports: a survey of patients undergoing cardiac surgery. JAMA. 1998;279:1638-42. [PMID: 9613914]

24. Van Herck P, De Smedt D, Annemans L, Rettmen R, Rosenthal MB, Sermeus W. Systematic review: Effects, design choices, and context of pay-for-performance in health care. BMC Health Serv Res. 2010;10:247. [PMID: 20731816]

25. Centers for Medicare & Medicaid Services. Decision memo for continuous positive airway pressure (CPAP) therapy for obstructive sleep apnea (OSA) (CAG-00093R). 2005. Accessed at www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=204&fromdb=true on 9 April 2012.

26. Bernstein SJ, Kosecoff J, Gray D, Hampton JR, Brook RH. The appropriateness of the use of cardiovascular procedures. British versus U.S. perspectives. Int J Technol Assess Health Care. 1993;9:3-10. [PMID: 8423114]

27. Gray D, Hampton JR, Bernstein SJ, Kosecoff J, Brook RH. Audit of coronary angiography and bypass surgery. Lancet. 1990;335:1317-20. [PMID: 1971385]

28. Jaspers MW, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc. 2011;18:327-34. [PMID: 21422100]

29. Kawamoto K, Del Fiol G, Strasberg HR, Hulse N, Curtis C, Gimino JJ, et al. Multi-National, Multi-Institutional Analysis of Clinical Decision Support Data Needs to Inform Development of the HL7 Virtual Medical Record Standard. AMIA Annu Symp Proc. 2010:2010:377-81. [PMID: 21347004]

30. Persell SD, Kaiser D, Dolan NC, Andrews B, Levi S, Khandekar J, et al. Changes in performance after implementation of a multifaceted electronic-health-record-based quality improvement system. Med Care. 2011;49:117-25. [PMID: 21178789]

31. Sequist TD, Zaslavsky AM, Colditz GA, Ayanian JZ. Electronic patient messages to promote colorectal cancer screening: a randomized controlled trial. Arch Intern Med. 2011;171:636-41. [PMID: 21149743]

32. Sistrom CL, Dang PA, Weilburg JB, Dreyer KJ, Rosenthal DI, Thrall JH. Effect of computerized order entry with integrated decision support on the growth of outpatient procedure volumes: seven-year time series analysis. Radiology. 2009;251:147-55. [PMID: 19221058]

33. Vartanians VM, Sistrom CL, Weilburg JB, Rosenthal DI, Thrall JH. Increasing the appropriateness of outpatient imaging: effects of a barrier to ordering low-yield examinations. Radiology. 2010;255:842-9. [PMID: 20501721]
Current Author Addresses: Dr. Baker: Feinberg School of Medicine, 750 North Lake Shore Drive, Chicago, IL 60611.
Dr. Qaseem: American College of Physicians, 190 N. Independence Mall West, Philadelphia, PA 19106.
Dr. Reynolds: University of Virginia, PO Box 800761, Charlottesville, VA 22908.
Dr. Gardner: ECRI Institute, 5200 Butler Pike, Plymouth Meeting, PA 19462-1298.
Dr. Schneider: RAND, 20 Park Plaza Boston, MA 02116.

Author Contributions: Conception and design: D.W. Baker, A. Qaseem, P.P. Reynolds, L.A. Gardner.
Analysis and interpretation of the data: A. Qaseem, P.P. Reynolds, L.A. Gardner.
Drafting of the article: D.W. Baker, A. Qaseem, L.A. Gardner.
Critical revision of the article for important intellectual content: A. Qaseem, P.P. Reynolds.
Final approval of the article: D.W. Baker, A. Qaseem, P.P. Reynolds, E.C. Schneider.
Collection and assembly of data: A. Qaseem, E.C. Schneider.