

Developing Quality Measures to Address Overuse

Jason S. Mathias, MD, MS

David W. Baker, MD, MPH

MEDICAL SOCIETIES AND OTHER ORGANIZATIONS have worked for many years to develop quality measures to assess underuse of beneficial health care services (eg, failure to use antiplatelet therapy for patients with acute myocardial infarction). More recently, organizations have begun efforts to address unnecessary tests and treatments and the high costs of health care by developing “overuse” measures (eTable, available at <http://www.jama.com>).¹⁻⁴ Overuse is defined as the use of a service that is unlikely to improve patient outcomes or for which potential harms exceed likely benefits.⁵

The standards for developing and evaluating overuse measures are not clearly defined and may differ from those for underuse measures.³ Two key issues need careful consideration in the development, implementation, and evaluation of overuse measures as compared with more typical underuse measures: level of evidence and mitigating potential unintended consequences.

Level of Evidence Required for Overuse Measures

Overuse measures should target services for which there is strong evidence (eg, high-quality randomized controlled trials for therapeutics or high-quality prospective cohort studies for diagnostic tests⁶) that a service is not beneficial. Ideally, there should also be a professional guideline that confirms that the available evidence is sufficient to recommend against the use of a service and that delineates the specific clinical circumstances in which the service should not be used and any important exceptions (ie, specific clinical situations in which use is justified).

However, unlike underuse measures, for which randomized controlled trial-level evidence is desired, there are good arguments to consider developing an overuse measure for a service, even if there is a lower level of evidence or lack of a guideline recommendation against the measure. Some diagnostic and therapeutic services are widely used, even for patients for whom no clear indication exists. For example, a Center for Medicare & Medicaid Services (CMS) overuse measure has identified frequent use of “double” thoracic computed tomography (CT) scanning—a rarely indicated study in which a CT scan with contrast and a CT scan without contrast are performed in series.⁴ Even if no direct evidence exists against a service, it is reasonable to undertake efforts to control its use if the service was never shown beneficial prior to widespread use and if the majority of experts doubt that the service is of value in a target population.

Conversely, good evidence may exist that a service is not beneficial, but guidelines may not have made explicit rec-

ommendations against the service because some guideline developers are less accustomed to issuing recommendations against services and because they may be concerned about negative reactions from physicians or the public if they do so. For example, the United States Preventive Services Task Force recommendation against screening for prostate cancer using prostate-specific antigen testing was heavily criticized.^{7,8} Thus, rather than relying only on randomized controlled trials and guidelines, it is necessary to assess the quality, quantity, and consistency of evidence supporting an overuse measure, including prospective cohort studies, retrospective observational studies using existing databases, and expert opinion obtained via a formalized consensus process.

The level and strength of evidence for overuse measures may need to vary depending on the potential harms and benefits of the service under consideration. For a service with significant potential harms (eg, chemotherapy), even the absence of proven efficacy may justify developing an overuse measure. For example, the American Society of Clinical Oncology recommends (with rare exceptions) against using “cancer-directed therapy for solid tumors when the patient has low performance status (3 or 4), has experienced no benefit from prior evidence-based interventions, and is ineligible for a clinical trial and *when there is no strong evidence supporting the clinical value of further anticancer treatment*” [emphasis added].² Developers of such measures should consider harms and costs (including total societal costs) from the cascade of testing or treatments that follow an initial service.

Mitigating Potential Unintended Consequences of Overuse Measures

Overuse measures may lead to underuse of beneficial diagnostic or therapeutic procedures.³ To mitigate this risk, overuse measures should be designed to ensure that patients for whom a service may improve outcomes are excluded from the measure denominator. The goal should be to maximize specificity (correctly labeling care as overuse) at the expense of sensitivity (completely identifying all cases of overuse). For example, the CMS Imaging in Low Back Pain measure excludes magnetic resonance imaging (MRI) scans performed in patients with “red flags” (eg, cancer, fever, weakness), either because MRI scans performed in these patients have potential benefits that outweigh potential risks

Author Affiliations: Division of General Internal Medicine (Dr Mathias) and Division of General Internal Medicine and Geriatrics, Institute for Public Health and Medicine (Dr Baker), Feinberg School of Medicine, Northwestern University, Chicago, Illinois. **Corresponding Author:** David W. Baker, MD, MPH, Division of General Internal Medicine and Geriatrics, Institute for Public Health and Medicine, Feinberg School of Medicine, Northwestern University, 750 N Lakeshore Dr, 10th Floor, Chicago, IL 60611 (dwbaker@northwestern.edu).

Table. Potential Unintended Consequences of Overuse Measurement

Potential Unintended Consequences of Overuse Measurement	Example
Underuse of the service when it is actually indicated	Measuring overuse of bone scanning for staging low-risk patients may unintentionally lead to underuse in higher-risk patients for whom a bone scan is indicated
Underuse of other related services	Measuring overuse of cervical cancer screening in low-risk women may unintentionally lead to underuse of other preventive services, such as screening mammography
Patient selection	Measuring overuse of spine imaging may unintentionally lead primary care physicians to avoid caring for patients with low back pain by referring them all to specialists
Care location shift	Measuring overuse of imaging in physicians' offices may unintentionally lead to physicians sending patients to the emergency room for imaging
Increasing use of alternate tests or treatments	Measuring overuse of lumbar spine radiographs may unintentionally lead to increased use of lumbar spine magnetic resonance imaging
Damage to the patient-physician relationship	Measuring overuse of antibiotics for bronchitis may unintentionally damage a physician's relationship with a patient because the physician did not order the treatment the patient desired
Clinician dissatisfaction with quality measurement	Excessive measurement burden (eg, additional documentation to confirm that service use is not overuse) may lead to clinician dissatisfaction with quality measurement on the front lines, gaming the system, or both, to improve performance on the measure without improving patient care
Adverse public health effects	Measuring overuse of blood cultures may unintentionally lead to decreased availability of data necessary to track antibiotic resistance over time; such difficult situations must be anticipated and addressed prior to implementation of overuse measures

or because studies that defined efficacy of MRI excluded these patients.⁴

Developers of measures should also ensure that the data used to identify exclusions are readily available, valid across multiple care settings, and accurate indicators of clinical characteristics that indicate that a service is beneficial. Failure to meet any of these criteria could potentially lead to patient harm. For example, patients taking anticoagulants should be excluded from a measure of overuse of brain CT scans after atraumatic headache. If accurate data on patients' medications are not available, it would be impossible to reliably differentiate patients for whom a CT scan may be beneficial (eg, those taking anticoagulants) from those for whom the potential harms outweigh potential benefits (eg, those not taking anticoagulants).

In addition, it will be important to routinely monitor for underuse of the service in situations in which it is indicated, underuse of other related services, or increasing use of alternate tests or treatments. For example, a measure of bone scan overuse for staging patients with prostate cancer at low risk of metastatic disease could unintentionally decrease bone scan use in patients at high risk for metastatic disease. To monitor for this potential unintended consequence, preimplementation and postimplementation underuse of bone scans in patients at high risk for bone metastases should be measured. Similarly, a measure of overuse of cervical cancer screening could unintentionally decrease the use of other related preventive services (eg, screening mammography).

The TABLE reports other potential unintended consequences. Each overuse measure may have a unique set of hypothetical unintended consequences. Developers of overuse measures should engage patients, physicians, and other stakeholders to help identify potential unintended consequences that are most likely to be problematic. Developers should then outline a strategy to monitor for unintended consequences if an overuse measure is implemented, especially as part of pay-for-performance programs. For difficult-to-monitor potential unintended consequences, it may be helpful to examine the effect of overuse measures via studies in a controlled environment (eg, a large

multispecialty group practice with a robust electronic health record system for tracking service use patterns and adverse outcomes) prior to widespread implementation.

Conclusions

Compared with traditional underuse measures, the rules of evidence for developing overuse measures are less well defined, and thoughtful strategies are needed to avoid unintended consequences of overuse measures. When carefully developed, implemented, and monitored, overuse measures have the potential to be part of the solution to the cost, quality, and safety problems in the US health care system.

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Baker reported receiving honoraria for research presentations from Abbott Laboratories (health literacy) and Merck (quality measurement) and currently receiving grant funding from the Agency for Healthcare Research and Quality and the National Institutes of Health. Dr Mathias reported no disclosures.

Additional Contributions: We would like to thank the members of the American Medical Association–convened Physician Consortium for Performance Improvement working group on overutilization measures for their contributions to the design and writing of this paper, including Helen Burstin, MD, MPH, Joseph Drozda, MD, Jeff Linder, MD, MPH, Mark Metersky, MD, and Jeremiah Schuur, MD, MHS. None of these persons received compensation for their contributions. We also thank the other members of the Measures Implementation and Evaluation Committee for their feedback.

Online-Only Material: The eTable is available at <http://www.jama.com>.

REFERENCES

- National Priorities Partnership. *National Priorities and Goals: Aligning Our Efforts to Transform America's Healthcare*. Washington, DC: National Quality Forum; 2008.
- Choosing Wisely website. <http://www.choosingwisely.org>. 2012. Accessed September 9, 2012.
- Baker DW, Qaseem A, Reynolds PP, Gardner LA, Schneider EC; American College of Physicians Performance Measurement Committee. Design and use of performance measures to decrease low-value services and achieve cost-conscious care. *Ann Intern Med*. 2013;158(1):55-59.
- Mathias JS, Feinglass J, Baker DW. Variations in US hospital performance on imaging-use measures. *Med Care*. 2012;50(9):808-814.
- Institute of Medicine, Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press; 2001.
- Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
- Volk RJ, Wolf AM. Grading the new US Preventive Services Task Force prostate cancer screening recommendation. *JAMA*. 2011;306(24):2715-2716.
- Miller DC, Hollenbeck BK. Missing the mark on prostate-specific antigen screening. *JAMA*. 2011;306(24):2719-2720.