

RONALD ADLER, MD, FAAFP, SAMI HAMDAN, MPH,
COURTNEY SCANLON, MD, AND WAYNE ALTMAN, MD, FAAFP

Quality Measures: How to Get Them Right

Pay-for-performance programs have produced disappointing results. Fewer and more appropriate, evidence-based quality measures could help.



The way we deliver and pay for health care in the United States has changed significantly in the last 50 years. The current emphasis on value-based care over traditional fee for service has led to the development of more than 2,500 quality measures¹ used to incentivize physicians and health care organizations to improve quality of care and reduce cost. The latest and most comprehensive effort to tie quality measurement to payment is the Quality Payment Program, which resulted from the Medicare Access and CHIP Reauthorization Act (MACRA). (See “Pay for performance and cost control: a brief history” with the online version of this article: <https://www.aafp.org/fpm/2018/0700/p23.html>.)

Despite the proliferation of quality measures and the pay-for-performance (P4P) systems that use them, there is little evidence of resulting positive changes in physician behavior or patient outcomes.²⁻⁴ Instead, most P4P systems have led to significant

ABOUT THE AUTHORS

Dr. Adler is an associate professor of family medicine and community health at the University of Massachusetts School of Medicine in Worcester, Mass., and co-founder of Care That Matters, a physician advocacy group. Sami Hamdan is a fourth-year medical student at Tufts University School of Medicine, Boston. Dr. Scanlon is a second-year resident at the Tufts Family Medicine Residency at Cambridge Health Alliance in Cambridge, Mass. Dr. Altman is a professor and the Jaharis Family Chair of Family Medicine at Tufts University School of Medicine. He is also a co-founder of Care That Matters. Author disclosures: no relevant financial affiliations disclosed.

administrative burdens⁵ and unintended consequences.⁶ Quality measures tied to financial incentives often crowd out the intrinsic motivation of physicians, par-

Most pay-for-performance systems have led to significant administrative burdens and unintended consequences.

ticularly for complex cognitive tasks,⁷ devaluing the patient-physician relationship and contributing to clinician burnout.

BUILDING A BETTER MEASURING STICK

As quality measures have rapidly increased, meeting and reporting the measures has become increasingly burdensome for health care providers, which has contributed substantially to rising health care costs.⁵ The bewildering complexity of public and private payment schemes, often using different measures and benchmarks to assess similar episodes or activities, has increased this burden for physicians.⁸

The Centers for Medicare & Medicaid Services, America's Health Insurance Plans (AHIP), and the American Academy of Family Physicians are part of a multi-stakeholder group working to harmonize and align measures across public and private payers. Reducing overlaps and gaps in metrics and reporting requirements is

essential to easing the costs and burdens associated with quality measurement. While it is clearly important to reduce the number of measures and improve the way they are implemented, the measures themselves should also be carefully scrutinized. Process measures with unclear clinical benefit should be replaced with *outcome* measures that are evidence-based and patient-centered. This effort should also recognize that many measures lack sufficient research support and were implemented without proper validation or vetting.⁶ Even the Medicare Quality Payment Program includes measures that fail to satisfy articulated criteria.^{9,10} (See "Quality measures: good and bad," page 25.)

To this end, representatives of Care That Matters, a physician group [including the authors] that advocates for improved quality measure design and implementation, and the editors of DynaMed Plus, an online evidence-based medicine reference, have created 10 criteria that can be used to assess the appropriateness of health care quality measures.

1. THE ASSESSED OUTCOME MATTERS TO PATIENTS

Measures should fundamentally address the things that matter most to patients: reducing mortality, improving quality of life, and lowering costs. Ideally, health care systems provide care that patients need and want. This requires eliciting and honoring patient preferences. Poorly designed P4P systems can create conflicting interests between clinicians and patients. For example, measures that focus strictly on a disease or surrogate outcomes may encourage physicians to pursue interventions that do not produce clinical improvements valued by patients. Disease-oriented measures such as A1C results are widely utilized but are often not ideal markers of a patient's health status.¹¹ Rather, evidence and consensus increasingly support outcome measures that are patient-centered, patient-reported, or both.¹²

2. THE PATIENT WILL EXPERIENCE A NET BENEFIT

There should be sufficient evidence that clinical actions associated with a measure lead to benefits that are likely to outweigh harms.

KEY POINTS

- The proliferation of health care quality measures and pay-for-performance programs has not led to significant improvements in patient outcomes but has contributed to greater administrative burdens for physicians.
- Some groups are working to consolidate the number of quality measures, especially measures with unclear benefits or a lack of evidentiary support.
- Quality measures should emphasize outcomes important to patients, provide them with a net benefit, and preserve their autonomy.
- Quality measures should also encourage behavior that leads to improved health, offer benefits that outweigh the resource expenditure, discourage "gaming," be specific, focus on outcomes the physician can influence, and consider social determinants of health.

QUALITY MEASURES: GOOD AND BAD

USEFUL

Tobacco Use: Screening and Cessation Intervention (National Quality Forum 0028)

Description: Percentage of patients 18 years and older who were screened for tobacco use at least once within 24 months and, if identified as a tobacco user, received tobacco cessation intervention.

Why it's good: Cessation of tobacco use is associated with decreased risk for heart disease, stroke, and lung disease – all outcomes that clearly matter to patients, and net benefit has been clearly established.¹ The resource requirements for screening, intervening, and reporting are all modest. This intervention clearly preserves patient autonomy, and it is reasonably resistant to “gaming.” The group of patients targeted (18 years and older; at least one preventive patient encounter or two patient encounters in the measurement period, excluding those with limited life expectancy) and the numerator, which (for tobacco users only) includes brief counseling, pharmacotherapy, or both. There are no significant issues related to physician control or social determinants of health.

PROBLEMATIC

Breast Cancer Screening (National Quality Forum 2372)

Description: Percentage of women ages 50–74 who had a mammogram to screen for breast cancer within 27 months prior to the end of the measurement period.

Why it's bad: By presuming that every woman in the specified age group needs and wants a mammogram, the measure does not respect patient autonomy, ignoring that many reasonable and well-informed women choose not to get one. Breast cancer screening involves trade-offs among potential benefits and harms, the rates of which tend to be respectively over- and under-estimated. For example, there is no all-cause mortality benefit, and the breast cancer-specific mortality benefit is small (1 per 1,299 women screened from age 50–59 years),² while more than 60 percent of women in their 50s will experience at least one false positive result,³ many then having unnecessary biopsies. This measure also fails to exclude women with limited life expectancy.

Physicians should offer breast cancer screening to eligible women and inform them of the potential benefits and harms. It would be more appropriate for this measure to assess physicians on whether they engage patients in shared decision making, not what patients decide, which is beyond the physician's control.

1. Patnode CD, Henderson JT, Thomson JH, Senger CA, Fortmann SP, Whitlock EP. Behavioral counseling and pharmacotherapy interventions for tobacco cessation in adults, including pregnant women: a review of reviews for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2015;163(8):608-621.

2. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 U.S. Preventive Services Task Force recommendation. *Ann Intern Med.* 2016;164(4):244-255.

3. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study [published correction appears in *Ann Intern Med.* 2014;160(9):658]. *Ann Intern Med.* 2011;155(8):481-492.

Nearly every health care intervention has the capacity to harm patients, so physicians should have strong – and easily accessible – scientific evidence that a particular quality measure will lead to a net positive health outcome for the patient. For instance, the “Choosing Wisely” campaign analyzes net benefits to discourage unnecessary and potentially harmful lab tests.¹³

Quality measures should also account for patients with multiple conditions.

Many current disease-oriented clinical guidelines were designed for individual conditions but may produce unintended harmful consequences for patients who have comorbidities.¹⁴

3. IMPLEMENTING THE MEASURE IMPROVES OUTCOMES

There should be sufficient evidence that implementation of the measure leads to benefits that outweigh harms. While

evidence for a test, treatment, or other intervention tied to a measure is necessary for it to be considered appropriate, it is insufficient. There can be negative unintended consequences of implementing a measure, even though the intervention itself may be evidence-based. Evidence should also demonstrate that use of the measure will not result in misuse of the test, treatment, or other intervention in

The physician whose quality of care is being measured should have sufficient authority, influence, or capacity to affect performance on the quality measure.

ways that lead to poorer health outcomes. Measure implementation should be shown to induce appropriate, evidence-based care.

4. IMPLEMENTING THE MEASURE USES AN APPROPRIATE AMOUNT OF RESOURCES

Implementation of the measure should produce net benefits that justify the resource (human, material, and financial) expenditure, including resources required for patient care, measurement, and reporting. Ideally, the benefits of the measure should outweigh the time and resources required to implement it.⁶ Quality measures do not necessarily require high administrative burdens or financial penalties to promote quality improvement and adherence to evidence-based practice.¹⁵ Each measure creates not only administrative costs but also potential opportunity costs because resources devoted to measure attention are not available for other interventions that might have a more positive impact on patients' health.⁵

5. THE PATIENT RETAINS HIS OR HER AUTONOMY

Many quality measures presume that there is a single, best approach to a given clinical situation, but this is not always the case. For example, some cancer screening tests produce many harms, including false positive results, unnecessary biopsies, and over-diagnosis of indolent cancers, even though there

is no all-cause mortality benefit. Given this trade-off between benefits and harms, physicians should pursue shared decision making with their patients to choose the option that best reflects the evidence and the patients' personal characteristics, values, and preferences. Once informed, patients may choose to pursue or defer screening. However, many prevailing quality measures actually reward clinicians for the number of screens they perform, not whether shared decision making occurred. Pressures from disease-oriented measures or process measures should not unduly limit patient autonomy in essential health care decisions.¹⁶

6. "GAMING" OR MANIPULATION IS NOT ENCOURAGED

A quality measure should not motivate a significant number of physicians to change their patient selection, clinical decision-making behavior, or reporting in ways that improve measure performance but not health outcomes. The risks and costs of so-called "gaming" in health care are regularly debated.¹⁷ Gaming is likely to persist or worsen as cost, administrative burden, and complexity of payment methods continue to increase. Common examples of gaming include altering reported data, manipulating diagnostic coding, and "cherry-picking," or selectively excluding the sickest or most challenging patients who would likely contribute to poor clinician performance on P4P measures.¹⁸

7. THE DENOMINATOR IS CLEARLY SPECIFIED

The population to whom the measure is applied must be clearly and adequately specified with appropriate exclusion criteria and assessment methods. When quality measures are applied across a population, there will be some patients for whom the measure is less suitable due to individual factors.⁶ Judicious use of exclusion criteria for specific subpopulations can mitigate this challenge.

8. THE NUMERATOR IS CLEARLY SPECIFIED

The desired outcome, test, treatment, or other intervention must be clearly described with criteria and a timeline for action, all supported by evidence.¹⁰

9. THE PHYSICIAN CAN INFLUENCE THE OUTCOME

The physician whose quality of care is being measured should have sufficient authority, influence, or capacity to affect performance on the measure and should not be penalized for factors beyond their control.⁶

In some situations, quality should be measured at a system level, incentivizing systems to provide resources and infrastructure that support physicians in providing high quality, team-based care.⁶ This is essential given that physicians work in multidisciplinary teams and are increasingly part of a hospital or accountable care organization. Further, measures that adopt a system focus may help promote care coordination among the many clinicians and other caregivers who interact with patients. A system focus could help improve other quality measures and diminish the fragmentation of health care.

10. SOCIAL DETERMINANTS OF HEALTH ARE CONSIDERED

Many P4P programs create a distinct disadvantage for physicians and health systems that care for vulnerable populations.¹⁸ Measures must acknowledge the limits of a physician's ability to influence an outcome, especially when results primarily reflect the patient's socioeconomic status. For example, if patients with hypertension cannot afford their prescriptions or patients with diabetes cannot access healthy food options, their physicians cannot easily or successfully improve patient health outcomes. Ideally, measures would also account for patient variability, particularly given the nuances of complex, chronic illnesses and comorbidities. Not only can such factors affect health status more than the quality of health care, they also can interfere with a physician's ability to achieve high performance on many quality measures. Systems for risk adjustment and risk stratification should be robust enough to accurately capture the variance in health caused by social determinants.¹⁹

NEXT STEPS

Most current quality measures are not supported by evidence that they promote outcomes that matter, such as reducing mortality, improving quality of life, or

WHAT CAN A DOCTOR DO?

There are several ways family physicians can facilitate solutions to problems with quality measures.

Advocate and educate. Discuss inappropriate quality measures and their use with your colleagues, your organization's leadership, and your local, state, and national medical societies. Look for opportunities to express your views more publicly, such as by writing an opinion piece for a newspaper. Prioritize discussing how quality measures can affect patients, especially in terms of their harms and costs. Acknowledge the need to measure quality while advocating for de-implementation of bad measures and for a more methodical, evidence-based approach to developing and implementing good ones. Emphasize the importance of measuring things that matter to patients.

Control what you can. If you are a leader in your health care plan, delivery system, or practice, try to influence the selection of measures for which family physicians will be accountable. Negotiate based on your understanding of which quality measures are appropriate. You may succeed in having some proposed measures excluded and others designated "for feedback only" and not used to affect compensation.

When you are coerced into activities that do not align with your values, the resulting dissonance can be very stressful and contribute to burnout. Prioritize the quality measures that you consider most meaningful for you and your patients. If, like most physicians, you have limited resources for population health management and quality improvement, you cannot optimize performance on all of them.

lowering costs. Inappropriate measures can induce harms, including wasteful overtreatment, adverse effects, distraction from more meaningful health care interventions, and acceleration of physician burnout. Meanwhile, there are fundamental problems with P4P programs that limit their utility, though it remains an open question whether P4P programs that use better measures could be more successful at producing intended results. (See "What can a doctor do?")

We suggest de-implementation of many health care quality measures until a new generation of evidence-based measures is developed and tested against predefined criteria for appropriateness, such as those presented above. Quality measures that have not been shown to promote improved, meaningful outcomes that matter to patients should not be used in P4P programs. **FPM**

1. National Quality Measures Clearinghouse. Agency for Healthcare Research and Quality. <https://qualitymeasures.ahrq.gov/browse/setting>. Accessed June 20, 2018.

2. Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy*. 2013;110(2-3):115-130.
3. Mendelson A, Kondo K, Damberg C, et al. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. *Ann Intern Med*. 2017;166(5):341-353.
4. Levine DM, Linder JA, Landon BE. The quality of outpatient care delivered to adults in the United States, 2002 to 2013. *JAMA Intern Med*. 2016;176(12):1778-1790.
5. Casalino LP, Gans D, Weber R, et al. U.S. physician practices spend more than \$15.4 billion annually to report quality measures. *Health Aff (Millwood)*. 2016;35(3):401-406.
6. Saver BG, Martin SA, Adler RN, et al. Care That Matters: quality measurement and health care. *PLoS Med*. 2015;12(11):e1001902.
7. Himmelstein DU, Ariely D, Woolhandler S. Pay-for-performance: toxic to quality? Insights from behavioral economics. *Int J Health Serv*. 2014;44(2):203-214.
8. Blumenthal D, Malphrus E, McGinnis JM, eds. Committee on Core Metrics for Better Health at Lower Cost, Institute of Medicine. *Vital Signs: Core Metrics for Health and Health Care Progress*. Washington, DC: National Academies Press, 2015.
9. MacLean CH, Kerr EA, Qaseem A. Time out – charting a path for improving performance measurement. *N Engl J Med*. 2018;378(19):1757-1761.
10. Drabkin A, Scanlon C, Adler R. Checking the checkboxes – an evidence based review of quality performance measures. Poster session presented at: 5th Annual Lown Institute Conference - Research Symposium. May 5-7, 2017; Boston.
11. Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. *BMJ*. 2011;343:d7995.
12. Frosch DL. Patient-reported outcomes as a measure of healthcare quality. *J Gen Intern Med*. 2015;30(10):1383-1384.
13. Wolfson D, Santa J, Slass L. Engaging physicians and consumers in conversations about treatment overuse and waste: a short history of the choosing wisely campaign. *Acad Med*. 2014;89(7):990-995.
14. Tinetti ME, Bogardus ST Jr, Agostini JV. Potential pitfalls of disease-specific guidelines for patients with multiple conditions. *N Engl J Med*. 2004;351(27):2870-2874.
15. Cassel CK, Jain SH. Assessing individual physician performance: does measurement suppress motivation? *JAMA*. 2012;307(24):2595-2596.
16. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. *JAMA*. 2004;291(19):2359-2366.
17. Woolhandler S, Ariely D, Himmelstein DU. Why pay for performance may be incompatible with quality improvement. *BMJ*. 2012;345:e5015.
18. Roberts ET, Zaslavsky AM, McWilliams JM. The value-based payment modifier: program outcomes and implications for disparities. *Ann Intern Med*. 2018;168(4):255-265.
19. Daniel H, Bornstein SS, Kane GC, Health and Public Policy Committee of the American College of Physicians. Addressing social determinants to improve patient care and promote health equity: an American College of Physicians position paper. *Ann Intern Med*. 2018;168(8):577-578.

Send comments to fpmedit@aafp.org, or add your comments to the article online.

THE AAFP'S LARGEST ANNUAL MEETING

AAFP
FMX
FAMILY MEDICINE
EXPERIENCE

New Orleans

OCTOBER 9-13, 2018 • AAFP.ORG/FMX

LIKE NO OTHER

REGISTER SOON TO SECURE YOUR SPOT AT THE LOWEST PRICE.

PAY FOR PERFORMANCE AND COST CONTROL: A BRIEF HISTORY

Medicare – 1965	Medicare was established to provide health insurance to those 65 years of age and older, covering both inpatient and outpatient services. As access increased and technology advanced, costs soared. Between 1967 and 1983, Medicare reimbursements to physicians and hospitals increased tenfold, ¹ inspiring a priority shift in the 1980s from access to cost containment.
Diagnosis Related Groups (DRGs) and Relative Value Units (RVUs) – 1983	DRGs were introduced to replace fee-for-service with a prospective payment system based on the average cost to deliver care for a specific "case." The complex formula was primarily designed to encourage inefficient hospitals to improve. In 1983, Medicare also introduced the Resource-Based Relative Value Scale, which pays physicians based on the number of RVUs assigned to services. RVUs are based on time spent, required skill and training of the physician, practice expenses, and malpractice expense. Over time, this system came to overvalue procedural services at the expense of cognitive services. ²⁻³
Health Maintenance Organizations (HMOs) and Capitation	HMOs continued cost containment efforts in the 1990s using capitation to pay physicians a flat rate for each assigned patient. "Cherry picking" healthier patients and undertreatment were unintended consequences of capitation. ⁴
Medicare Sustainable Growth Rate (SGR) – 1997	The SGR was enacted to control Medicare spending on physician services, tying payments to inflation. ⁵ Full implementation would have reduced physician payments annually. Congress intervened 17 times, preventing this with the so-called "doc fix," although the annual uncertainty created instability. It was repealed in 2015 as part of MACRA legislation.
Publication of <i>To Err is Human</i> by the Institute of Medicine (IOM) – 1999	The origins of pay for performance developed in part from a desire to control rapidly rising health care costs and address gaps in quality of care. The IOM's <i>To Err is Human</i> report inspired widespread efforts to optimize patient safety. ⁶ Meanwhile it was reported that health care in the United States ranked lower than in many other developed countries. ⁷
Affordable Care Act (ACA) – 2010	The ACA expanded the use of metrics in health care. It included incentives to increase care coordination and population health management through accountable care organizations – coordinated networks of physicians and hospitals primarily designed to improve care and reduce cost.
Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)	The Quality Payment Program (QPP) of MACRA accelerates the transition from fee-for-service to value-based payment. Under MACRA, physicians' Medicare payments will be based on quality of care and other performance measures. For more details on MACRA and the QPP, see "Making Sense of MACRA in 2018: Six Things You Need to Know," <i>FPM</i> , January/February 2018. https://www.aafp.org/fpm/2018/0100/p21.html .

1. Gottlob P, ed. *Medicare Hospital Prospective Payment System: How DRG Rates Are Calculated and Updated*. Washington, D.C.: Department of Health and Human Services; 2001. <https://oig.hhs.gov/oei/reports/oei-09-00-00200.pdf>.

2. Berenson RA, Goodson JD. Finding value in unexpected places – fixing the Medicare physician fee schedule. *N Engl J Med*. 2016;374(14):1306-1309.

3. Laugesen MJ. The resource-based relative value scale and physician reimbursement policy. *Chest*. 2014;146(5):1413-1419.

4. Gosden T, Forland F, Kristiansen IS, et al. Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians. *Cochrane Database Syst Rev*. 2000;(3):CD002215.

5. Hirsch JA, Harvey HB, Barr RM, et al. Sustainable growth rate repealed, MACRA revealed: historical context and analysis of recent changes in Medicare physician payment methodologies. *AJNR Am J Neuroradiol*. 2016;37(2):210-214.

6. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. *Qual Saf Health Care*. 2004;13(2):145-151.

7. The World Health Organization. *The World Health Report 2000. Health Systems: Improving Performance*. Geneva: World Health Organization, 2000. <http://www.who.int/whr/2000/en/>. Accessed June 15, 2018.