## Editorial

# Measuring Value in Primary Care: Enhancing Quality or Checking the Box?

Over the last three decades, numerous health care process measures and surrogate health outcome measures have exploded onto the scene, often driven by payers wanting to show commitment to "value." With increasing popularity of pay-for-performance and other value-based incentive programs aimed at maximizing quality while minimizing costs, primary care practices are confronted with an ever-increasing sea of quality metrics that they are urged to satisfy. In our health system at the University of Michigan, Blue Cross Blue Shield—a payer that spends one in five medical claim dollars on value-based payment arrangements—incentivizes at least 200 different quality metrics (Abelson 2014). Primary care physicians (PCPs) constantly chase this dizzying array of metrics, recognizing that such measures frequently fail to help them care for their sickest, most vulnerable patients. When seeing a homeless patient with uncontrolled diabetes and food insecurity, for example, recording their smoking history seems less pressing. Time spent at a practice on such "quality metrics" diverts from addressing the highest priority areas for the patient—areas that often do not have a quality checkbox from a payer. PCPs face this tension daily in balancing "checkbox" care that supposedly achieves high "quality" with the actual clinical value they strive to provide vulnerable patients.

Aligning health care payments with quality measurement has long been promoted as a vehicle for achieving higher quality of care (2001), but which existing metrics, if any, truly measure quality? For example, does the mere documentation of smoking status equate to quality? What if the patient does not want to follow the advice of the physician? Is the physician providing bad quality care? Should the physician be paid less? If the goal is ultimately to improve health, quality must be measured as meaningful health improvement, recognizing the role patients play in the interaction, and incentivized accordingly.

Measuring and reporting quality has been a challenge as quality scores were first publicly reported for surgeons performing coronary artery bypass grafting (CABG) in New York. Despite widespread enthusiasm for this type of quality measurement and reporting, there is concern that such programs may actually penalize physicians caring for medically and psychosocially complex patients (Hofer et al. 1999). In the case of CABG ratings, physicians may have been incentivized to "dump" their sicker patients (Werner and Asch 2005). Similarly, PCPs may be willing to fire patients who do not comply with their medical recommendations (Farber et al. 2008). This dumping behavior is actually encouraged by incentive programs that promote compliance with simple quality metrics but fail to account for the complex interventions necessary for care of high-risk patients. The unintended consequence may be decreased access to care, lower quality of care—or both—for many vulnerable patients with multiple chronic conditions or difficult socioeconomic circumstances (McMahon, Hofer, and Hayward 2007). Thus, the challenge health services researchers face today is determining how best to use performance measures to incentivize health care quality for those who need it most.

In this issue of *Health Services Research*, Jason Wang and colleagues offer a notable case study of three overlapping incentive programs promoting quality improvement in a network of small primary care practices in New York City. They examine the impact of meaningful use, patient-centered medical home (PCMH) and pay-for-performance incentives on seven quality metrics comprising both process and outcome measures related to smoking, obesity, blood pressure, blood sugar, and vascular disease management. Practices participating in all three programs demonstrated improvement in the study's chosen quality metrics during the follow-up period between 2009 and 2012, with PCMH practices achieving improvement on the greatest number of measures. However, each program chose different quality measures to incentivize, leading to potential complexity and confusion during implementation in individual practices. While there was observed improvement in some clinically relevant metrics such as delivery of smoking cessation interventions, other "quality improvements" were largely driven by process measures that do not directly impact patient

Address correspondence to Renuka Tipirneni, M.D., M.Sc., Robert Wood Johnson Foundation Clinical Scholar, Clinical Lecturer in Internal Medicine, University of Michigan, 2800 Plymouth Rd, Bldg 10, Rm G016, Ann Arbor, MI 48109-2800; e-mail: rtipirne@med.umich.edu. Laurence F. McMahon, Jr., M.D., M.P.H., is with the Division of General Medicine, Department of Internal Medicine; Institute for Healthcare Policy and Innovation; Department of Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor, MI.

outcomes, such as the simple documentation of smoking status and body mass index. Focusing on identifying which incentive programs most effectively drive improvements in quality metrics overlooks a more fundamental issue—how do we meaningfully measure quality in primary care?

Health services researchers have measured quality at various levels, but commonly in ways that are not clinically relevant. First, we have recorded whether a health care process is done or not done—these checklist-like process measures are not nuanced, but are relatively easy to measure and improve (often by increasing documentation alone). Second, we have evaluated health outcome measures. Given the long follow-up required to obtain meaningful outcomes such as disability and mortality, we have more often assessed surrogate measures of health, such as hemoglobin A1C for diabetes control. These intermediate outcomes often employ arbitrary cut-offs and rarely take into account which patients would benefit most from chronic disease management such as blood sugar lowering in diabetes (Kerr et al. 2003). For example, a diabetic patient with a high baseline hemoglobin A1C (severe disease) may benefit more from aggressive blood sugar lowering, even if he or she never meets an arbitrary cut-off of a hemoglobin A1C less than 7–8 (recommended diabetic control) (Vijan, Hofer, and Hayward 1997). Setting an arbitrary goal blood glucose target for the whole population may disincentivize primary care providers from caring for high-risk patients who would benefit most from disease management. Even what we once thought were the best measures to incentivize—for example, low density lipoprotein—often turn out to be the wrong target and have to be revisited (Hayward et al. 2010; Smith and Grundy 2014). Thus, it is unclear whether either process or outcome measures target the right goal (Kerr et al. 2001). To improve population health, we need more clinically nuanced measurements of quality that weight the value of high-priority care more and low-priority care less (McMahon and Heisler 2008).

It is time to abandon the one-size-fits-all approach that we know does not work. Instead we need to advance quality improvement metrics to the next level by using population-based modeling to incentivize approaches aimed at achieving the greatest health improvements in primary care practices, as determined by patients and their primary care providers. We could utilize the vast amount of "big data" available in health systems to decide which patients to target and how to optimally improve health outcomes in a given population. This would entail use of multivariable risk assessment and stratification to identify high-risk subgroups of patients that are most likely to benefit from certain interventions (McMahon et al. 2005). Practices could tailor quality improvement programs to their local population and in-

centivize what is most valuable to improving their patients' health. Few populations could benefit more from such health improvement than low-income individuals with Medicaid or medically complex patients. Many Medicaid managed care plans have actually employed such modeling to target high-risk patients for care management programs (2008). Other plans and health systems may similarly use the wealth of data available to them to create tailored programs for their high-risk patients.

As Wang et al. mention in their article, recording of risk factors is a "critical first step in prevention of adverse outcomes." However, this one step is not sufficient to improve quality in primary care. With advances in predictive modeling and increasing availability of electronic health information, we can do much more to meaningfully improve care delivery. We need to take the next step of population-based risk stratification and tailored interventions that provide true clinical value to vulnerable patients who need care most. It is time to stop incentivizing checkbox medicine and to start incentivizing health.

## ACKNOWLEDGMENTS

Renuka Tipirneni
Laurence F. McMahon, Jr.

## REFERENCES

Arora, R., J. Boehm, L. Chimento, L. Moldawer, and C. Tsien. 2008. *Designing and Implementing Medicaid Disease and Care Management Programs: A User's Guide.* Rockville, MD: Agency for Healthcare Research and Quality.

Abelson, R. 2014. "Health Insurers Are Trying New Payment Models, Study Shows." *New York Times* B3, July 9, 2014.

Committee on Quality of Health Care in America, Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: 2001 by the National Academy of Sciences.

Farber, N. J., M. E. Jordan, J. Silverstein, V. U. Collier, J. Weiner, and E. G. Boyer. 2008. "Primary Care Physicians' Decisions about Discharging Patients from Their Practices." *Journal of General Internal Medicine* 23 (3): 283–7.

Hayward, R. A., H. M. Krumholz, D. M. Zulman, J. W. Timbie, and S. Vijan. 2010. "Optimizing Statin Treatment for Primary Prevention of Coronary Artery Disease." *Annals of Internal Medicine* 152 (2): 69–77.

Hofer, T. P., R. A. Hayward, S. Greenfield, E. H. Wagner, S. H. Kaplan, and W. G. Manning. 1999. "The Unreliability of Individual Physician "Report Cards" for Assessing the Costs and Quality of Care of a Chronic Disease." *Journal of the American Medical Association* 281 (22): 2098–105.

Kerr, E. A., S. L. Krein, S. Vijan, T. P. Hofer, and R. A. Hayward. 2001. "Avoiding Pitfalls in Chronic Disease Quality Measurement: A Case for the Next Generation of Technical Quality Measures." *American Journal of Managed Care* 7 (11): 1033–43.

Kerr, E. A., D. M. Smith, M. M. Hogan, T. P. Hofer, S. L. Krein, M. Bermann, and R. A. Hayward. 2003. "Building a Better Quality Measure: Are Some Patients with 'Poor Quality' Actually Getting Good Care?" *Medical Care* 41 (10): 1173–82.

McMahon, L. F., Jr., and M. Heisler. 2008. "Looking for Quality under the Lamppost." *Journal of General Internal Medicine* 23 (11): 1923–5.

McMahon, L. F., Jr., T. P. Hofer, and R. A. Hayward. 2007. "Physician-Level P4P–DOA? Can Quality-Based Payment Be Resuscitated?" *American Journal of Managed Care* 13 (5): 233–6.

McMahon, L. F., Jr., R. Hayward, S. Saint, M. E. Chernew, and A. M. Fendrick. 2005. "Univariate Solutions in a Multivariate World: Can We Afford to Practice as in the "Good Old Days"?" *American Journal of Managed Care* 11 (8): 473–6.

Smith, S. C., Jr., and S. M. Grundy. 2014. "2013 ACC/AHA Guideline Recommends Fixed-Dose Strategies Instead of Targeted Goals to Lower Blood Cholesterol." *Journal of the American College of Cardiology* 64 (6): 601–12.

Vijan, S., T. P. Hofer, and R. A. Hayward. 1997. "Estimated Benefits of Glycemic Control in Microvascular Complications in Type 2 Diabetes." *Annals of Internal Medicine* 127 (9): 788–95.

Wang, J. J., J. Cha, K. M. Sebek, C. M. Mccullough, A. S. Parsons, J. Singer, and S. C. Shih. 2014. "Factors Related to Clinical Quality Improvement for Small Practices Using an EHR." *Health Services Research* 49 (6): 1729–1746.

Werner, R. M., and D. A. Asch. 2005. "The Unintended Consequences of Publicly Reporting Quality Information." *Journal of the American Medical Association* 293 (10): 1239–44.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.