

EDITORIALS

Improving the Performance of Performance Measurement

LeChauncy D. Woodard, MD, MPH^{1,2}, and Laura A. Petersen, MD, MPH^{1,2}

¹Health Policy and Quality Program, Houston VA HSR&D Center of Excellence, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA; ²Section of Health Services Research, Baylor College of Medicine, Houston, TX, USA.

J Gen Intern Med 25(2):100-1

DOI: 10.1007/s11606-009-1198-z

© Society of General Internal Medicine 2009

Performance measurement programs are increasingly being used to determine physician compensation and to inform consumer decisions about their health care.^{1,2} Although performance measurement and pay-for-performance programs³ hold great promise for improving quality, combating physician skepticism about the motivation for and accuracy of such programs is vital to their success.^{4,5} While physicians overwhelmingly believe that financial incentives should be given for high quality care, fewer than one-third think that current performance measures are accurate, and only slightly more feel that those responsible for designing quality measures will work to ensure their accuracy.⁶ It is understandable that in the era of evidence-based medicine, physicians expect rigorous statistical methods and approaches for performance measurement that are reproducible and robust. Thus, failure to design methodologically rigorous performance measurement programs may limit physician buy-in and hinder quality improvement.

Despite similarities in the dimensions of care being measured, the methodologies employed to assess performance can vary significantly, thereby producing inconsistent results^{3,4} and creating concern among stakeholders. In order to fully engage physicians and to achieve the intended aim of improving quality, it is imperative that these performance measurement initiatives utilize valid and transparent methods. Imprecise measurement may lead to unintended consequences, including erroneously identifying physicians as poor performers, physician avoidance of seriously ill patients that may negatively impact ratings, and consumer choice based on flawed data.² Therefore, it is reasonable for stakeholders to expect accurate and robust measures of performance.

In this issue of the *Journal of General Internal Medicine*, Landon et al. address a critical first step in accurate performance measurement—appropriate attribution of patients to the providers responsible for their care.⁷ The authors note that, although frequently cited as a shortcoming of current performance measurement efforts, little empiric evidence exists to support the assertion that faulty attribution of patients can lead to differences in quality. Landon and colleagues convincingly demonstrate that apparent discrepancies in quality may instead reflect variations in methods for attributing patients to providers. Using administrative data, the authors employed three different sampling strategies

ranging from most inclusive (one visit over a 3-year time period) to least inclusive (one visit during the study year and at least one visit during the prior 2 years) to determine eligible patients for cancer screening at nine community health centers. Cancer screening rates were then varied across groups, with the highest screening rate assigned to patients identified using the most stringent methods, thus demonstrating some evidence of continuity, and the lowest rates assigned to patients identified using the least stringent methods and therefore having the lowest levels of continuity of care. They found that choice of sampling strategy can significantly alter the distribution of eligible patients at community health centers, thereby resulting in wide variations in overall performance at some facilities, particularly those with less stable patient populations. These findings provide compelling evidence to support the importance of defining the appropriate denominator to ensure accurate quality assessment.

What can we generalize from this work to ongoing provider profiling and performance measurement work? Because the alternate methods used to attribute patients to providers inherently reflect varying levels of continuity, the influence on type of measure being assessed, process or outcome, will likely differ as well. For example, process measures, such as ordering a test (e.g., hemoglobin A1c level), could be easily achieved with a single encounter. Conversely, outcome measures (e.g., blood pressure control) may require multiple visits involving several medication adjustments and counseling regarding lifestyle modifications to attain.⁸⁻¹⁰ Thus, less stringent sampling methods may unfairly penalize providers who care for the most complex, chronically ill patients, highlighting the importance of correct patient attribution in this population.

The authors focus on performance measurement for preventive care practices noting that, regardless of algorithm used to assign patients, large numbers of individuals were eligible for each of the screening tests at all facilities. However, if similar sampling strategies were applied to determine eligible patients for measures of chronic illness care, one would expect different results. Hofer and colleagues assessed the reliability of performance measures for diabetes, a highly prevalent chronic condition, and found that physicians would require more than 100 diabetic patients to accurately detect variation in care. Yet the vast majority of physicians had less than 60 eligible patients in their panels, limiting the ability to make valid comparisons and increasing the likelihood that performance will be disproportionately influenced by outliers.¹¹ To address this important problem, we are currently examining the use of methods developed to profile academic institutions to assess the performance of providers with smaller samples of chronically ill patients.¹² Nonetheless, even if these methods

can be successfully implemented using fewer patients, this does not diminish the importance of accurately assigning patients to the providers responsible for their care, as demonstrated by Landon and colleagues.

The work by Landon and colleagues adds to a growing list of issues that should be addressed to improve the science of performance measurement.^{8-10,13,14} We have shown that diabetic patients with life-limiting chronic conditions are less likely to achieve hemoglobin A1c and low-density lipoprotein control despite more frequent monitoring. While intense glycemic and lipid control is likely to confer little benefit to this patient population, patients with these conditions are rarely excluded from performance assessments and focus on such measures diverts attention from other aspects of care that may be more pertinent to these patients.¹³ Other issues that should be addressed include the types of patient-specific data necessary to create measures that more accurately reflect quality of care, benefit to patient, and patient preferences. Also, creation of composite measures of performance as well as measures reflecting episodes of care will be important advances.¹⁵

Lastly, in addition to improving the science of performance measurement, it seems reasonable to expect that ideal performance measures should inform health-care providers about how they can improve their care. While we have demonstrated the feasibility of collecting measures that incorporate such actionable information within the VA health-care system,⁸ widespread implementation may need to await more sophisticated medical record systems in non-VA health-care settings. Although performance measurement and its uses continue to progress, there is little doubt that such initiatives will become even more ubiquitous as we embark on efforts to reform our current health-care system. We need more work in this important and challenging area of health services research.

Acknowledgement: *The authors thank Cassie Robinson, MPH, for assistance with the literature review for this editorial. This work is supported in part by VA HSR&D PPO 09-316-1 (PI LeChauncy D. Woodard, MD, MPH), VA HSR&D IIR 04-349 (PI Laura A. Petersen, MD, MPH), NIH R01 HL079173-01 (PI Laura A. Petersen, MD, MPH), and Houston VA HSR&D Center of Excellence HFP90-020 (PI Laura A. Petersen, MD, MPH). Dr. Petersen is an American Heart Association Established Investigator Awardee (grant no. 0540043N).*

Disclosure: *The views expressed are solely of the authors and do not necessarily represent those of the VA. There are no conflicts of interest to disclose.*

Corresponding Author: *LeChauncy D. Woodard, MD, MPH; Health Policy and Quality Program, Houston VA HSR&D Center of Excellence, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA (e-mail: lwoodard@bcm.tmc.edu).*

REFERENCES

1. **Rosenthal MB, Dudley RA.** Pay-for-performance: will the latest payment trend improve care? *JAMA.* 2007;297:740-4.
2. **Werner RM, Asch DA.** The unintended consequences of publically reporting quality information. *JAMA.* 2005;293:1239-44.
3. **Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S.** Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006;145:265-72.
4. **Draper DA.** Physician performance measurement: a key to higher quality and lower cost growth or a lost opportunity? Commentary No. 3. Center for Studying Health System Change, Washington DC (June 2009).
5. **Hayward RA, Kent DM.** 6 EZ steps to improving your performance: (or How to Make P4P Pay 4U!). *JAMA.* 2008;300:255-6.
6. **Casalino LP, Alexander GC, Jin L, Konetzka RT.** General internists' views on pay-for-performance and public reporting of quality scores: a national survey. *Health Aff.* 2007;26:492-9.
7. **Landon BE, O'Malley AJ, Keegan T.** Can choice of the sample population affect perceived performance: implications for performance assessment. *J Gen Intern Med.* 2010.
8. **Petersen LA, Woodard LD, Henderson LM, Urech TH, Pietz K.** Will hypertension performance measures used for pay-for-performance programs penalize those who care for medically complex patients? *Circulation.* 2009;119:2978-85.
9. **Kerr EA, Smith DM, Hogan MM, Hofer TP, Krein SL, Bermann M, Hayward RA.** Building a better quality measure: are some patients with 'poor quality' actually getting good care? *Med Care.* 2003;41:1173-82.
10. **Selby JV, Uratsu CS, Fireman B, Schmittziel JA, Peng T, Rodondi N, Karter AJ, Kerr EA.** Treatment intensification and risk factor control: toward more clinically relevant quality measures. *Med Care.* 2009;47:395-402.
11. **Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG.** The unreliability of individual physician "Report Cards" for assessing the costs and quality of care of a chronic disease. *JAMA.* 1999;281:2098-105.
12. **Draper D, Gittoes M.** Statistical analysis of performance indicators in UK higher education. *J R Statist Soc A.* 2004;167(Part 3):449-74.
13. **Woodard LD, Urech T, Robinson C, Kuebler M, Pietz K, Petersen LA.** Differences in therapy intensification for glycemic and lipid control in diabetic patients with and without limited life expectancy. *J Gen Intern Med.* 2009;24(S1):S55-6.
14. **Woodard L, Urech T, Kuebler M, Pietz K, Capistrano L, Petersen LA.** The impact of comorbid medical conditions on the quality of diabetes care. *J Gen Intern Med.* 2007;22(S1):138.
15. **Profit J, Gould JB, Zupancic JA, Pietz K, Petersen LA.** Selecting quality metrics for a composite index of NICU quality. *E-PAS* 2009;5510.176.