has taken the lead in focusing on patient care outcomes and attaching financial incentives to those providers, hospitals, and nursing homes achieving good results, or avoiding certain poor results or complications.

The fact that there remain so many points and counterpoints regarding quality reporting,[3] and that such reporting is already in use for determining physician and other practitioner payments should, I believe, compel us all to do our part to improve these systems, so that we all find ourselves happily on the "point" side of the argument in the not-too-distant future.

To read or post commentaries in response to this article, see it online at http://www.AnnFamMed.org/content/15/3/204.

Key words: patient care outcomes; health care quality assessment; assessment, outcomes (health care)

Submitted March 20, 2017; accepted March 21, 2017.

References

1. Frankel BA, Bishop TF. A cross-sectional assessment of the quality of physician quality reporting system measures. *J Gen Intern Med.* 2016;31(8):840-845.
2. 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.
3. Caveney BJ. Pay-for-performance incentives: holy grail or sippy cup? *N C Med J.* 2016;77(4):265-268.

POINT/COUNTERPOINT

Counterpoint: How Quality Reporting Made Me a Worse Doctor

David L. Hahn, MD, MS[1,2]

[1]Wisconsin Research and Education Network (WREN), Madison, Wisconsin

[2]Department of Family Medicine and Community Health, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin

Ann Fam Med 2017;15:206-208. https://doi.org/10.1370/afm.2077.

The current approach to accountability of medical care is to blend reporting of "quality measures" with "pay-for-performance" (P4P).[1] Benefits of this approach include use of medical evidence and population-based thinking. Limitations include use of disease-oriented instead of patient-oriented measures, and arbitrary benchmarks lacking actionable information. Evidence that physician P4P strategies have improved patient care and outcomes is limited.[1] Pay-for-performance incentives to maximize performance instead of incentivizing informed patient preferences can put clinicians in the position of having to choose between providing excellent individualized patient care, or being paid equitably. Linking compensation with achieving arbitrary benchmarks conflicts with practicing shared decision making wherein the quality measure is the adequacy of the shared–decision-making encounter, not the prevalence of the eventual outcome chosen by the patient.[2] These perverse incentives made me a worse doctor as indicated by failing to meet the benchmarks.

Conflicts of interest: author reports none.

CORRESPONDING AUTHOR

David L. Hahn, MD, MS
Department of Family Medicine and Community Health
University of Wisconsin School of Medicine and Public Health
1100 Delaplaine Court
Madison, WI 53715
dlhahn@wisc.edu

PATIENT SATISFACTION

My partners complained about the conflict between good medical practice versus giving patients what they demanded (such as unneeded antibiotics and/or opioids) to increase patient satisfaction scores. System factors beyond the direct control of the clinician may also demoralize clinicians who feel they are being unfairly judged.[3] In one study, whether patients chose (higher satisfaction) or were assigned (lower satisfaction) their doctor was 10 times more influential than clinician behavior.[3] Might one also expect an inverse association between patient satisfaction scores and open access scheduling? I asked myself that question as I continued to keep my practice open to "work-ins," "walk-ins," and new patients.

MEASURES BASED ON OPINION NOT EVIDENCE

A measure based on expert opinion increased the death rate of the unlucky patient group that achieved the "quality" benchmark.[4] It took 4 years from the time of publication of a randomized controlled trial—(Action to Control Cardiovascular Risk in Diabetes [ACCORD])[4] that provided Level 1 evidence that aggressive treatment of Type 2 diabetes to an A_{1c} below 7 increased mortality—before this measure was eliminated from my practice.[5] During those 4 years we who were aware of ACCORD had to choose between evidence-based practice and looking good on the "quality" measure. Here's another example, though not as deadly. There is no strong evidence for mass screening of depression because all existing trials are of high-risk groups (eg, high utilizers of medical care[6] or groups containing patients with previous depression diagnoses).[7,8] Depression measures should encourage case finding.[9] In my experience, mass screening wastes resources.

PATIENT-ORIENTED MEASURES

Advocates for disease-oriented measures (eg, $HgbA_{1c}$) argue that surrogates (like laboratory values) are easier to measure than patient-centered outcomes (such as morbidity and mortality). Is it better to do the inadequate thing systematically than back off and do the right thing whenever possible? A growing number of validated patient-reported outcome measures are available (eg, PHQ-9 for depression, the Asthma Control Test for asthma). Process measures for depression (visit frequency) and asthma (medication use) are of limited utility[10] or valueless.[11] Only surrogate measures validated in practice-based research effectiveness randomized controlled trials should be considered for use. Directly measuring patient-important outcomes is better.

ARBITRARY BENCHMARKS

Current quality measures reward achieving a high prevalence of performance and rank performance based on arbitrary benchmarks (eg, the benchmark for systolic blood pressure [SBP] control is <140). Is achieving a SBP of 138 better than a SBP of 142? This approach inevitably leads to gaming.[12] Performance measures should draw attention to clinical conditions that most warrant attention (eg, treating a SBP of 220 to 150 is clinically important but does not meet the current benchmark). My partners and I knew this and felt helpless to do anything about it. The original intent of quality measurement was to inform valid quality improvement activities. Benchmarks need to be reconfigured to fulfill this aim.

SHARED DECISION MAKING

Shared decision making is a process in which the clinician offers options to her patient who is encouraged to apply his own values to making the choice that is best suited to him.[13,14] Shared decision making is appropriate for clinical preventive services and management of chronic conditions that form the bulk of current primary care practice. Quality assessment should focus on the shared decision making process, not on the prevalence of the choices made by the patient.[2] A clear conflict of interest exists for clinicians practicing in settings that link achievement of arbitrary benchmarks to clinician pay or other incentives/disincentives. This may be the most disturbing unanticipated consequence of the "quality" movement. I was an early advocate of clinical preventive service delivery in primary care[15-17] and knew I could manufacture high numbers if I wanted to.[18] I refused to play the game, however, because I had learned that shared decision making was more personally rewarding. This inevitably meant that "quality" reporting made me look like a worse doctor.

My partners complained that we were "not making widgets." I wonder to what extent clinician burnout may be attributable to knowing that one is being judged unfairly by metrics that undermine effective practice.[12] Measures must be improved. They should provide actionable information. They should align with good clinical practices and promote patient-centered care, especially shared decision making. They should encourage reflection and valid continuous quality improvement. They should undergo regular evaluation and should allow for changes in response to data and provider input.[1] Measures should not be used to arbitrarily and spuriously reward or punish clinicians. Current "quality" measures do not address many things that stakeholders (patients, clinicians, payers) feel are important.[19] The Institute of Medicine has recently outlined a radically different set of core quality measures.[20] How many years (or decades) before we see better measures?

To read or post commentaries in response to this article, see it online at http://www.AnnFamMed.org/content/15/3/206.

Key words: quality improvement; quality measures; shared decision making

Submitted November 23, 2016; submitted, revised, January 13, 2017; accepted February 8, 2017.

References

1. Kondo KK, Damberg CL, Mendelson A, et al. Implementation processes and pay for performance in healthcare: a systematic review. *J Gen Intern Med.* 2016;31(Suppl 1):61-69.
2. Hahn DL. Public reporting needs reform! *J Fam Pract.* 2009;58(5): 237-238, 240.

3. Schmittdiel J, Selby JV, Grumbach K, Quesenberry CP Jr. Choice of a personal physician and patient satisfaction in a health maintenance organization. *JAMA*. 1997;278(19):1596-1599.
4. Gerstein HC, Miller ME, Byington RP, et al; Action to control cardiovascular risk in diabetes study group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358(24):2545-2559.
5. Wisconsin-Collaborative-for-Healthcare-Quality-(WCHQ). http://www.wchq.org/. Accessed Jun 4, 2016.
6. Pearson SD, Katzelnick DJ, Simon GE, Manning WG, Helstad CP, Henk HJ. Depression among high utilizers of medical care. *J Gen Intern Med*. 1999;14(8):461-468.
7. Thombs BD, Arthurs E, El-Baalbaki G, Meijer A, Ziegelstein RC, Steele RJ. Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review. *BMJ*. 2011;343:d4825.
8. Thombs BD, Ziegelstein RC, Roseman M, Kloda LA, Ioannidis JP. There are no randomized controlled trials that support the United States Preventive Services Task Force Guideline on screening for depression in primary care: a systematic review. *BMC Med.* 2014;12(1):13.
9. Hahn DL. How practice-based research changed the way I manage depression. *J Fam Pract*. 2003;52(10):784-788.
10. Rost K, Dickinson LM, Fortney J, Westfall J, Hermann RC. Clinical improvement associated with conformance to HEDIS-based depression care. *Ment Health Serv Res*. 2005;7(2):103-112.
11. Crans Yoon A, Crawford W, Sheikh J, Nakahiro R, Gong A, Schatz M. The HEDIS Medication Management for People with Asthma Measure is Not Related to Improved Asthma Outcomes. *J Allergy Clin Immunol Pract*. 2015;3(4):547-552.
12. Lowe T, Wilson R. Playing the game of outcomes-based performance management. Is Gamesmanship inevitable? Evidence from theory and practice. [published online ahead of print]. *Soc Policy Adm*. doi 10.1111/spol.12205.
13. Légaré F, Ratté S, Stacey D, et al. Interventions for improving the adoption of shared decision making by healthcare professionals. *Cochrane Database Syst Rev*. 2010;(5):CD006732.
14. Stacey D, Légaré F, Col NF, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2014;1(1):CD001431.
15. Hahn DL. Feasibility of sigmoidoscopic screening for bowel cancer in a primary care setting. *J Am Board Fam Pract*. 1989;2(1):25-29.
16. Hahn DL. Systematic cholesterol screening during acute care visits. *J Am Board Fam Pract*. 1993;6(6):529-536.
17. Hahn DL, Olson N. The delivery of clinical preventive services: acute care intervention. *J Fam Pract*. 1999;48(10):785-789.
18. Hahn DL, Berger MG. Implementation of a systematic health maintenance protocol in a private practice. *J Fam Pract*. 1990;31(5):492-502, discussion 502-504.
19. Etz RS, Gonzalez MM, Brooks EM, Stange KC. Less AND more are needed to assess primary care. *J Am Board Fam Med*. 2017;30(1):13-15.
20. Blumenthal D, McGinnis JM. Measuring Vital Signs: an IOM report on core metrics for health and health care progress. *JAMA*. 2015;313(19):1901-1902.