

All-or-Nothing Treatment Targets Make Bad Performance Measures

Rodney A. Hayward, MD

Last year, the National Committee for Quality Assurance (NCQA) approved 2 new Health Plan Employer Data and Information Set (HEDIS) measures for people with diabetes mellitus (DM): to achieve a glycosylated hemoglobin (A1C) level measurement of <7% and a blood pressure (BP) of <130/80 mm Hg. This decision was the culmination of a 5-year battle between DM advocates/experts and evidence-based medicine advocates/experts. In this issue of the *Journal*, Pogach et al* address the importance of considering comorbidities when implementing optimal care measures.¹ This editorial, however, will discuss the broader question of why this intuitively appealing approach—using “optimal” treatment goals as performance measures—will almost always require more sophisticated measurement approaches (like those proposed by Pogach and others) or else risk generating performance measures that are inaccurate, promote waste, and perhaps cause substantial patient harm.¹⁻⁷

The newly adopted performance measures are outcome measures, not processes, and like most outcomes, patients’ risk of having the outcome (A1C <7% or BP <130/80) will vary dramatically depending on patient attributes (disease severity, personal choices, human capital, and response to treatment). Even when risk adjustment is superb, outcome measures can be heavily prone to inaccuracy, statistical inefficiency, and vulnerability to gaming,^{3-4,6-8} which is why modern performance measurement efforts usually focus on processes (the care people receive). Therefore, readers may be perplexed as to why 2 new outcome measures lacking any risk adjustment were adopted. In truth, these new measures were a compromise between advocates of optimal goals (disease advocates) and advocates of simple, inexpensive performance measures (health plan leadership). Experts in medical evidence were not included in the compromise, which is part of the problem. It would be wonderful if measuring optimal care was simple and easy (and on rare occasions it is). In general, however, simple and easy measures are restricted to measuring poor care. Measuring more optimal care almost always requires added complexity and detail and particular attention to patient preferences and the risks and costs of the interventions needed to achieve idealized treatment goals.

Payers, disease advocates, consumer groups, and political leaders are often dismissive of the complex reality of measuring care, suggesting that performance measures may have problems, but medical-evidence experts have unrealistic standards and are supplying the enemies of accountability with ammunition. I am not unsympathetic to such concerns. There are enemies of performance measurement who nitpick at every little imperfection in performance measurements. I firmly believe that worrying about modest problems with measures is counterproductive. However, it is also true that wishful thinking will not transform poor performance measures into useful ones, and that well-meaning people have a profound aptitude for letting their desires and ideology blind them to unwanted facts and complexities that are so vexingly common in the real world. In particular, many leaders in performance measurement have little appreciation for, or interest in, how unimportant small deviations from “optimal” goals can be, or how treatment benefits and safety can vary widely across a patient population in complex ways, or that optimal care by definition must consider patient and societal preferences.^{2,9-13} Most of all, promoting optimal care using performance measures requires considering the very real tensions among treatment-related benefits and treatment-related burdens, risks, and costs. HL Mencken once said, “For every problem, there is a solution that is simple, neat, and wrong,”¹⁴ and using unadjusted “all-or-nothing” optimal treatment targets as performance measures is such an example.

Almost everyone who voted for the 2 new HEDIS measures did not know that the Technical Expert Panel (TEP) of the Diabetes Alliance¹⁴ unanimously rejected the proposed A1C <7% and BP <130/80 measures even after being asked to reconsider their decision by political leaders of the Alliance. The TEP, citing the evidence-based guidelines of the American College of Physicians and Veterans’ Affairs/Department of Defense, noted that the proposed performance measures

In this issue

*Related paper (Pogach et al) / p133

www.ajmc.com

Full text and PDF

All-or-Nothing Treatment Targets

were inconsistent with the medical evidence and that they had major concerns regarding perverse incentives inherent in the proposed measures. However, the TEP offered multiple compromise approaches to measuring the quality of glycemic and BP control, with each proposal rejected either because it was not strict enough (opposed by diabetes advocates) or was too complex (opposed by the health plans).

Blood Pressure Control

Often, the majority of patients who achieve an “optimal” treatment goal, like BP <130/80, are those with no or mild disease, meaning that the measure largely captures the disease severity of the patient population, and even worse, is often ineffective in capturing true variations in quality, especially the variations in the care of high-risk patients where most preventable morbidity and mortality resides.⁹⁻¹³ Dichotomizing the intermediate outcome measure makes this problem much worse. For example, using BP <130/80 as a thin, bright line results in rating the care of a patient with a naturally low BP as good care, but does a poor job of distinguishing truly good care from truly bad care. Even after prescribing 3 or 4 antihypertensive medications and paying careful attention to medication adherence, those with severe hypertension will usually have persistent elevations of systolic BP.¹³ The irony is that the BP <130/80 measure provides greater rewards for speculatively treating patients with mild disease (no clinical trial has demonstrated aggressively treating people with DM with mild BP elevations is either beneficial or safe) and does a poor job of rewarding the treatments shown in clinical trials to produce dramatic reductions in disability and mortality, because only a small minority of the severely hypertensive patients studied in the clinical trials achieved a systolic blood pressure (SBP) <130.¹³ Not only is there no evidence that using more than 3 or 4 medications in pursuit of the <130/80 goal is beneficial, but there is consistent grade B evidence that such treatment may increase cardiovascular mortality in those who have already achieved a diastolic blood pressure (DBP) <70.¹³⁻¹⁶ This is not a rare event, a quick look at National Health and Nutrition Examination Survey data reveals that of patients 65 years of age and older, about a third of people with DM with SBP >130 already have a DBP <70, meaning that the new HEDIS measure will frequently be promoting care that the best available evidence suggests will increase cardiovascular mortality.

Kerr et al have proposed measures that focus on how clinicians and systems respond to clinical indications (such as elevations in BP, A1C, and lipids) and how this improves accuracy and reduces perverse incentives for overtreatment and gaming.^{3,4} Such an approach would also result in avoiding

the safety concerns outlined above. However, nonclinicians tend to argue that prescribing the appropriate treatment is less important than whether the treatment target is achieved. This point of view shows a lack of understanding of epidemiological evidence and has been discussed fully elsewhere.^{9-13,17} Much of this literature, however, can be summed up as follows: (1) clinical trials assess treatments, not treatment targets; (2) in most instances, the risks associated with deviations from optimal goals follow an exponential pattern (so that small deviations from treatment targets are often of little importance); and (3) because most medical treatments have side effects, risks, and costs, pursuing minimal deviations from idealized goals will often involve speculative care that is not worth the cost and may expose patients to undue risks and burden.

A1C Control

The A1C <7% measure produces the same problems as those discussed above. In addition, the A1C <7% measure allows the manufacturers of several new and expensive hypoglycemic medications to promote their products using the A1C <7% measure as “evidence” that achieving this goal is very important. This point was not lost on the pharmaceutical industry, which funded a national campaign to get the A1C <7% measure adopted by the NCQA. This was a very small investment compared with the costs of funding clinical trials to evaluate whether these expensive new medications have benefits that are worth their risks (forget about cost-effectiveness; we do not even know if they are safe in pursuit of tight glycemic control). Although diabetes experts often point to the United Kingdom Prospective Diabetes Study (UKPDS) as evidence that achieving an A1C <7% is worth the unknown long-term risks of new medications, experts in epidemiological evidence will point out that the UKPDS shows no such thing.^{17,18} The UKPDS found that combination oral hypoglycemic therapy (metformin plus sulfonylurea) was associated with increased diabetes-related mortality, metformin monotherapy resulted in substantial patient benefit, and monotherapy with sulfonylureas or insulin had no significant effects on any of the prespecified patient outcomes, although the control achieved for A1C was very similar between groups.¹⁸ Further, although modeling studies strongly suggest that those with early-onset diabetes will get moderate benefit in the long term (if they live more than 15-20 years and you assume that the treatments have almost zero risk and disutility), older patients are very unlikely to get benefit.^{12,17} Once again, focusing on the dichotomized “optimal” treatment goal results in giving much more credit for providing speculative treatments for patients with mild disease than for providing high-priority

care. For example, a system that targets the new, expensive hypoglycemic treatments to patients who least need them (eg, patients with an A1C of 6.7%-7.7%) will do much better than one that targets these treatments to the best candidates for these treatments (eg, patients with an A1C of 8%-9%). This is why the TEP rejected the A1C <7% measure but proposed an A1C <8% or a weighted measure (one that places more importance on large deviations from optimal than trivial deviations) as alternatives.²⁻⁴

It is certainly a reasonable goal to want performance measures for optimal care standards to be simple, but optimal care is almost never simple. Some leaders in performance measurement have asked me, "Do you really think that these measures will lead clinicians and health systems to overtreat?" I am frankly amazed by this question. Spiraling healthcare costs and overtreatment are probably the defining features of the US healthcare system. Industry-funded "experts" and disease advocates have been effectively promoting overtreatment for decades, and performance measurement was supposed to be a tool to bring better value to healthcare spending. Although performance measurement has proved to be a very powerful tool,¹⁹ like all tools it provides opportunities for both benefit and harm. It is simply magical thinking to believe that performance measures will do good regardless of how haphazardly they are constructed, and that they will not do harm even when the measures adopted provide strong incentives for overtreatment.

Author Affiliation: From the VA Ann Arbor Health Services Research & Development, Center of Excellence, University of Michigan.

Correspondence Author: Rodney A. Hayward, MD, Director, VA Ann Arbor Health Services Research & Development, Center of Excellence, Professor of Medicine and Public Health, University of Michigan, 6312 Med Sci I, Ann Arbor, MI 48109-0604; E-mail: rhayward@umich.edu.

The opinions and views presented in this editorial are solely those of the author and do not necessarily represent those of the Department of Veterans Affairs or the University of Michigan.

REFERENCES

- Pogach LM, Tiwari A, Maney M, Rajan M, Miller DR, Aron D.** Should mitigating comorbidities be considered in assessing healthcare plan performance in achieving optimal glycemic control. *Am J Manag Care.* 2007;13:133-140.
- Pogach LM, Rajan M, Aron DC.** Comparison of weighted performance measurement and dichotomous thresholds for glycemic control in the Veterans Health Administration. *Diabetes Care.* 2006;29:241-246.
- Kerr EA, Krein SL, Vijan S, Hofer TP, Hayward RA.** Avoiding pitfalls in chronic disease quality measurement: a case for the next generation of technical quality measures. *Am J Manag Care.* 2001;7:1033-1043.
- Kerr EA, Smith DM, Hogan MH, et al.** Building a better quality measure: are some patients with "poor quality" actually getting good care? *Med Care.* 2003;41:1173-1182.
- O'Malley AS, Clancy C, Thompson J, Korabathina R, Meyer GS.** Clinical practice guidelines and performance indicators as related—but often misunderstood—tools. *Jt Comm J Qual Saf.* 2004;30:163-171.
- Hayward RA, Hofer TP, Kerr EA, Krein SL.** Quality improvement initiatives: issues in moving from diabetes guidelines to policy. *Diabetes Care.* 2004;27(suppl 2):B54-B60.
- 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;281:2098-2105.
- Rosenberg AL, Hofer TP, Strachan C, Watts CM, Hayward RA.** Accepting critically ill transfer patients: adverse effect on a referral center's outcome and benchmark measures. *Ann Intern Med.* 2003;138:882-890.
- Hayward RA, Hofer TP, Vijan S.** Narrative review: lack of evidence for recommended low-density lipoprotein treatment targets: a solvable problem. *Ann Intern Med.* 2006;145:520-530.
- Hayward RA, Kent DM, Vijan S, Hofer TP.** Reporting clinical trial results to inform clinical providers, payers and consumers. *Health Aff (Millwood).* 2005;24:1571-1581.
- McMahon LF, Hayward RA, Saint S, Chernew ME, Fendrick AM.** Univariate solutions in a multivariate world: can we afford to practice as in the "good old days"? *Am J Manag Care.* 2005;11:473-476.
- Vijan S, Hofer TP, Hayward RA.** Estimated benefits of glycemic control in microvascular complications in type 2 diabetes. *Ann Intern Med.* 1997;127:837-839.
- Vijan S, Hayward RA.** Treatment of hypertension in type 2 diabetes mellitus: blood pressure goals, choice of agents, and setting priorities in diabetes care. *Ann Intern Med.* 2003;138:593-602.
- H. L. Mencken Quotes.** Available at: <http://www.brainyquote.com/quotes/h/hlmencke162005.html>. Accessed February 20, 2007.
- National Diabetes Quality Improvement Alliance.** Available at: <http://www.nationaldiabetesalliance.org/organizations.html>. Accessed February 9, 2007.
- Messerli FH, Giuseppe Mancina G, Conti CR, et al.** Dogma disputed: can aggressively lowering blood pressure in hypertensive patients with coronary artery disease be dangerous? *Ann Intern Med.* 2006;144:884-893.
- Vijan S, Kent DM, Hayward RA.** Are randomized controlled trials sufficient evidence to guide clinical practice in type II (non-insulin dependent) diabetes mellitus? *Diabetologia.* 2000;43:125-130.
- Shaughnessy AF, Slawson DC.** What happened to the valid POEMs? A survey of review articles on the treatment of type 2 diabetes. *BMJ.* 2003;327:266.
- Asch SM, McGlynn EA, Hogan MM, et al.** Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. *Ann Intern Med.* 2004;141:938-945. ■