EDITORIAL AND COMMENT

The Next Generation of Clinical Performance Measures

David Atkins, MD, MPH

Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs, Washington, DC, USA.

J Gen Intern Med 31(Suppl 1):S3-5

DOI: 10.1007/s11606-015-3575-0

© Society of General Internal Medicine 2016

Many are familiar with the adage that “you can’t improve what you can’t measure.” Adopting clinical performance measures was an important part of the VA transformation of the 1990s, and measuring clinical performance has become an integral part of current efforts to drive improvement under Medicare, the Affordable Care Act (ACA), and in programs run by private payers. Measuring clinical performance, however, is a health system intervention, and as such, needs to be carefully examined. This includes how measures are created, how they are implemented, and the evidence for their potential benefits and harms. The landscape of healthcare has also changed substantially since the advent of the performance measurement movement, and it is time to update our approach to measuring quality. Rather than setting a single standard, identifying a random number of patients to whom the measure should apply, and conducting manual chart reviews at a single point in time, we need to be assessing quality for the entire population, including the more complex patients. A world with electronic health records and a much richer array of administrative and clinical data allows us to update our approach to include measures that consider population health, episodes of care, changes over time, and individual circumstances. The Department of Veterans Affairs (VA), which operates the largest integrated healthcare system in the United States, is a key contributor to the emerging knowledge base regarding how best to create and implement performance measures.

Funded by the VA’s Health Services Research and Development (HSR&D) Service, this *Journal of General Internal Medicine* (JGIM) Supplement is the product of an HSR&D-sponsored state-of-the-art (SOTA) conference titled “Next-Generation Clinical Performance Measures: Patient-Centered, Clinically Meaningful, and High Value,” held in 2014. After the conference, a Call for Papers was issued, and 14 manuscripts were submitted to JGIM. After rigorous JGIM peer review, ten articles were accepted for publication in this special JGIM Supplement. Published papers discuss empirical research on the effects of performance measurement on improvements in clinical care, as well as on unintended outcomes (e.g., inappropriate treatment or over-treatment). Papers also describe new methods and methodological

challenges in the selection and creation of performance measures that incorporate measures of benefit and harm, value, or patient preferences, as well as research on the implementation of performance measures that address human factors, incentives and facilitators, barriers, and expected and unintended consequences.

USING VA’S ELECTRONIC HEALTH RECORD SYSTEM TO MEASURE QUALITY

Three papers examined the ability to use VA electronic data to measure clinical quality in specific clinical areas. Phipps and colleagues developed and validated inpatient stroke electronic clinical quality measures that are part of the Meaningful Use (MU) program and VA efforts to improve inpatient stroke care.¹ The authors found that stroke MU indicators can be accurately generated from existing VA data in its electronic health record (EHR) system (nearly a 90 % match to chart review), but accuracy decreased slightly when data from the VA’s national-level corporate data warehouse (CDW) was used rather than more complete local data sources. The authors also found that a relatively small number of error types are responsible for a large number of the observed mismatches, suggesting specific areas in which EHR developers and informaticians could improve the accuracy of the electronic record for generating MU measures. The promise of these findings is that, with some tweaks and improvements, the accuracy of CDW data could improve to an extent sufficient to enable its accuracy to rival the more laborious process of collecting data region by region.

The VA has invested substantially in evidence-based mental health care, but the lack of accepted electronic performance measures for assessing depression treatment has made it difficult to rigorously evaluate its depression initiatives. Farmer and colleagues developed electronic population-based longitudinal quality metrics for depression care.² They found that despite rapid growth in the primary care VA population from FY 2000 to FY 2010, increasing by over one million veterans, the detection of new episodes of depression (8 %) and minimally appropriate treatment rates (84 %) remained stable. This suggests that the VA was able to maintain a standard of care while treating significantly more patients each year. At the same time, the authors caution that if the full spectrum of care—from detection to follow-up and treatment—is not captured, performance measures could actually mask the clinical areas in need of quality improvement.

Saini and colleagues discuss the development of an electronic measure of the overuse of screening colonoscopy for veterans who were screened between 2011 and 2013.³ They found that, compared to results obtained from manual record review, the electronic measure was highly specific (97 %) for overuse, but was not sensitive (20 %). To some extent, the fact that the electronic measure had low sensitivity reflects limitations in the ability to electronically ascertain screening indication. For example, it may be particularly challenging to differentiate the use of colonoscopy for screening from a non-screening indication (e.g., diagnostic purposes or post-polypectomy surveillance). The authors suggest that overuse measures could be combined with underuse measures to improve the appropriateness of colorectal cancer screening.

PERFORMANCE MEASURES FOR IMPROVING PATIENT-CENTERED CARE

Improving patient experience is a critical priority for the VA and for many other healthcare systems. Patient-reported experience measures (PREMs) are useful for assessing healthcare quality and safety, as well as patients' perceptions of care. Etingen and colleagues assessed the relationship between PREMs and healthcare quality metrics in a large sample of veterans receiving VA healthcare.⁴ Findings suggest that PREMs assessing elements of patient-provider communication (e.g., empathic provider care, shared decision-making) are most strongly associated with clinical indicators of chronic care management, while those relating more broadly to healthcare (e.g., patient activation, chronic illness care) are mainly related to measures of appropriate healthcare use (e.g., receiving appropriate preventive care, avoiding preventable hospitalizations and ER visits). Overall, PREMs are an effective way to engage patients, to consider their experiences and preferences, and to obtain a deeper understanding than that provided by simple measures of patient satisfaction.

PERFORMANCE MEASURES THAT ENHANCE VALUE-RELATED RESEARCH

Conversations with U.S. healthcare systems these days are as likely to refer to "value" as they are to "quality," since payers need to ensure that quality comes at a reasonable cost. The U.S. healthcare system increasingly embraces value measurement, both for reimbursement and quality improvement. To this end, an expert panel on value was convened during the VA HSR&D SOTA. Wagner and colleagues highlight findings from the panel on how to measure value, and they make recommendations for future research.⁵ Defining value as the incremental outcomes gained per dollar spent, the consensus of the panel was that understanding how to incentivize high-value care is a top priority for future research.

Pay-for-performance (P4P) programs use financial incentives to stimulate improvements in healthcare efficiency and quality.

However, P4P programs are complex interventions and vary widely in their implementation (i.e., characteristics of measures chosen, cost/efficiency, and incentive targets). Kondo and colleagues conducted a systematic review and key informant (KI) interviews to better understand how implementation features influence the effectiveness of P4P programs.⁶ While there was limited evidence from which to draw strong conclusions, findings suggest that P4P programs should evolve over time in response to periodic evaluation, and that they should target areas of poor performance. Moreover, measures and incentives should align with organizational priorities, and providers should be engaged in designing the implementation of measures.

In 2003, the VA tied a portion of executive compensation to performance on a quality measure meant to capture continuity of care for substance use disorder (SUD). At that time, performance-based bonuses could amount to 10 % of VA network directors' annual salaries; incentives were also provided to managers and clinicians at the discretion of the network directors. Using data for veterans with SUD who were treated in VA facilities from FY 2000 to FY 2009, Harris and colleagues evaluated whether implementing the measure resulted in expected improvements in performance.⁷ Findings show that including an SUD continuity-of-care quality measure in network directors' performance contracts was associated with an increase in measured performance, from 23 % just before the measure was implemented to 48 % by the end of the observation period. At the same time, the overall proportion of patients with SUD program contact that qualified for the measure decreased more rapidly over time following implementation, and varied significantly among facilities. The authors note that they were unable to determine what proportion of these changes were due to the desired improvements in patient management versus undesirable process changes to exclude patients from the measure's denominator. They propose a number of ways performance measures might be revised to reduce potential "gaming" of measures.

Other articles in this supplement cover an array of important topics related to performance measures. Prentice and colleagues focus on the need for specification and predictive validation of performance metrics, and discuss the VA's leadership in developing performance metrics through a planned upgrade in its electronic medical record system to enable the collection of more comprehensive VA and non-VA data.⁸ Hysong and colleagues evaluated the difficulty of completing clinical work associated with outpatient clinical performance measures.⁹ Finney and colleagues explore the relationship between a hospital's performance on measures of providing recommended care and patient outcomes.¹⁰ Because the process-outcome relationship is complex, they recommend multi-level analyses to evaluate process measure-outcome relationships at both the patient and hospital levels.

VA HSR&D is proud to have been an early and continued supporter of research on performance measurement and its role in improving quality of care. As we use an increasing number of measures to hold clinicians accountable for their clinical care, it

is just as important that we subject our measures—and the process for designing and implementing them—to the same level of scrutiny. We think these collected articles will make a significant contribution to our knowledge in this area.

Corresponding Author: David Atkins, MD, MPH; Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs, 810 Vermont Ave., NW (10P9H), Washington, DC, 20420, USA (e-mail: david.atkins@va.gov).

REFERENCES

1. Phipps MS, Fahner J, Sager D, et al. Validation of stroke meaningful use measures in a national electronic health record system. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3562-5.
2. Farmer MM, Rubenstein LV, Sherbourne CD, et al. Depression quality of care: measuring quality over time using VA electronic medical record data. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3563-4.
3. Saini S, Powell AA, Dominitz JA, et al. Developing and testing an electronic measure of screening colonoscopy overuse in a large integrated healthcare system. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3569-y.
4. Etingen B, Miskevics S, LaVela SL, et al. Assessing the associations of patient-reported perceptions of patient-centered care as supplemental measures of health care quality in VA. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3557-2.
5. Wagner TH, Burstin H, Frakt AB. Opportunities to enhance value-related research in the U.S. Department of Veterans Affairs. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3538-5.
6. Kondo KK, Damberg CL, Mendelson A, et al. Implementation processes and pay for performance in healthcare: a systematic review. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3567-0.
7. Harris AHS, Chen C, Rubinsky AD. Are improvements in measured performance driven by better treatment or “denominator management”? *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3558-1.
8. Prentice JC, Frakt AB, Pizer SD. Metrics that matter. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3559-0.
9. Hysong SL, Amspoker AL, Petersen LA. A novel method for assessing task complexity in outpatient clinical-performance measures. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3568-z.
10. Finney JW, Humphreys K, Kivlahan DR, Harris AHS. Excellent patient care processes in poor hospitals? Why hospital-level and patient-level care quality-outcome relationships can differ. *J Gen Intern Med*. 2015. doi:10.1007/s11606-015-3564-3.