

## PERSPECTIVE

## Metrics That Matter

*Julia C. Prentice, PhD<sup>1,2,3</sup>, Austin B. Frakt, PhD<sup>1,2,3,4</sup>, and Steven D. Pizer, PhD<sup>1,5</sup>*

<sup>1</sup>Health Care Financing and Economics, VA Boston Healthcare System, Boston, MA, USA; <sup>2</sup>Boston University School of Medicine, Boston, MA, USA; <sup>3</sup>Boston University School of Public Health, Boston, MA, USA; <sup>4</sup>Harvard T.H. Chan School of Public Health, Boston, MA, USA; <sup>5</sup>Northeastern University, Boston, MA, USA.

Increasingly, performance metrics are seen as key components for accurately measuring and improving health care value. Disappointment in the ability of chosen metrics to meet these goals is exemplified in a recent Institute of Medicine report that argues for a consensus-building process to determine a simplified set of reliable metrics. Overall health care goals should be defined and then metrics to measure these goals should be considered. If appropriate data for the identified goals are not available, they should be developed. We use examples from our work in the Veterans Health Administration (VHA) on validating waiting time and mental health metrics to highlight other key issues for metric selection and implementation. First, we focus on the need for specification and predictive validation of metrics. Second, we discuss strategies to maintain the fidelity of the data used in performance metrics over time. These strategies include using appropriate incentives and data sources, using composite metrics, and ongoing monitoring. Finally, we discuss the VA's leadership in developing performance metrics through a planned upgrade in its electronic medical record system to collect more comprehensive VHA and non-VHA data, increasing the ability to comprehensively measure outcomes.

**KEY WORDS:** metric validation; metric implementation; access to care metrics; mental health metrics; veterans.

J Gen Intern Med 31(Suppl 1):S70–3

DOI: 10.1007/s11606-015-3559-0

© Society of General Internal Medicine 2015

Enhancing value—i.e., improving outcomes as efficiently as possible—has gained salience across health systems. To maximize value, it must be measured.<sup>1</sup> The complexity of accurately measuring value has led to a proliferation of quality and performance metrics, often leading to confusion, duplication, and provider reporting burden.<sup>2,3</sup> To address these issues, a recent report from The Institute of Medicine (IOM) recommends stakeholders participate in a consensus-building process that identifies key health care goals focusing on outcomes that enhance value.<sup>2,3</sup> Once the key goals have been identified, metrics that measure these goals can be developed. Often, the IOM could not find existing measures that comprehensively captured the desired goals, resulting in a call for the development of comprehensive data and metrics in areas with identified gaps.<sup>2,3</sup>

We expand upon the IOM's findings on the need to simplify metrics through a consensus-building process that identifies core comprehensive metrics. Our experience working with performance metrics in the Veterans Health Administration (VHA) highlights other equally important challenges in the metric selection process. Key lessons discussed below include: specification and predictive validation, attention to implementation challenges to ensure data fidelity, and the ongoing evolution of data to comprehensively measure outcomes.

## VALIDATION OF METRICS

The IOM recommended that health systems measure 15 core metrics, including access to care. This is a domain in which the VHA has extensive experience. For over 15 years, Congress has requested that the VHA respond to concerns that visit waiting times are too long.<sup>4</sup> These concerns peaked in 2014 with accusations that delays in access to VHA care resulted in preventable deaths.<sup>5</sup> Throughout this long history, the VHA has employed a variety of strategies to measure and report on waiting times, learning some important lessons about the feasibility of validating and implementing performance metrics along the way.

The IOM committee argued that the best measure of access to care is unmet care need, assessed by surveys such as the Medical Expenditure Panel Survey or the VHA's Survey of Healthcare Experiences of Patients (SHEP).<sup>6,7</sup> Collection of survey data is costly and difficult to obtain in real time because of the effort required to contact patients and obtain their responses. Survey sampling frames also raise concerns of selection bias.<sup>8–10</sup> For example, only veterans who have used VHA services are surveyed in SHEP, limiting information about access for non-VHA users.<sup>7</sup> The ability to find and measure the entire population of patients that suffer from unmet health care needs in a timely matter is an important example of gaps in the metric landscape that the IOM identified.

Recognizing the limitations of survey-based access measures, our work has validated access metrics that can be constructed with routinely collected administrative data. This validation work has focused on both specification and predictive validity. Even conceptually compelling metrics can be useless in practice if they fail to measure the desired concept because the data underlying the metrics are misunderstood or inaccurate (i.e., specification validity).<sup>11</sup> Demonstrating specification validity, our work shows that several types of waiting

time metrics (e.g., number of days until next available appointment, number of days between appointment request and completed appointment) predict veterans' self-reported access. The relationship is seen across multiple outcomes self-reported in SHEP including patients' ability to obtain timely appointments, access to treatments and specialists, and overall satisfaction with VHA care.<sup>8</sup>

Predictive validation relates a potential performance metric to a conceptually related outcome in historical data.<sup>12</sup> If the metric is associated with the outcome as expected, users can have greater confidence that the metric truly reflects the intended concept. For example, if hospitals rated most highly for infection control processes also had high rates of post-surgical infections after controlling for patient characteristics, we might question the validity of the rating metric. In our work the same access metrics that demonstrated specification validity also demonstrated predictive validity by predicting clinical outcomes such as glycosylated hemoglobin levels, preventable hospitalization and mortality. These clinical outcomes are not likely to be related to metrics that fail to measure access and unmet health care needs.<sup>8,13-18</sup>

Two key lessons emerged from this validation work. First, different access metrics predicted satisfaction for different patient populations. The best metric for new patients (based on the date the patient first contacted the clinic) is not the same as the best metric for returning patients (based on the date of service requested by the patient). These population differences may not be clear in the IOM recommended consensus-building process that determines the original metrics, highlighting the need for validation. Second, when predicting patient-level outcomes for validation purposes, appropriate methods are needed to control for confounders such as patient-level health status. Without taking these confounders into account, true relationships between metrics and key outcomes can be obscured (e.g., longer waits predict improved health instead of poorer health). A common strategy to address this problem is to compute facility-level average metrics that exclude the individual patient whose outcome is being assessed.<sup>8,13-18</sup> This is a similar approach to using an instrumental variable analysis that allows consistent estimation of relationships between explanatory variables and outcomes even when key explanatory variables are correlated with the error terms in the models.

## MAINTAINING DATA FIDELITY OF METRICS

Even metrics based on initially appropriate and accurate data can fail over time if incentives to appear successful affect the fidelity of data generation or metric definition. The IOM recommends assessing the threats to metric validity that incentives can create.<sup>2,3</sup> We agree and argue that strong implementation design requires policymakers to anticipate potential unintended strategic responses from those whose performance

is being measured and design metrics to minimize the likelihood of such responses.

The VHA access crisis in 2014 highlighted the importance of this point and how it differs from validation. Our waiting time studies validated routinely collected access metrics by confirming that they predicted a wide variety of patient outcomes. Upon adoption of the metrics, the provision of bonuses to senior management for meeting access metric targets led to widespread concerns that access metrics were being manipulated and reported VHA waits from some facilities did not reflect actual patient experience.<sup>19-21</sup> The controversy resulted in the resignation of several senior officials including former Secretary Shinseki, the passage of the Veterans Access Choice and Accountability Act, and the current reorganization of the VHA that focuses on veteran-centric outcomes.<sup>5,22-25</sup>

The VHA experience that performance incentives can encourage behavior that can bias metrics is not unique. Roth et al. (2012) examined the effect of the HEDIS quality metric to avoid antibiotic prescribing for acute bronchitis on overall antibiotic prescribing. Antibiotic prescribing for acute bronchitis decreased sharply. However, overall antibiotic prescribing only decreased slightly because of a significant shift in diagnosis coding from acute to unspecified bronchitis.<sup>26</sup>

One strategy to reduce this danger is to use information from independent data sources. For example, as a measure of hospital quality, post-hospitalization mortality rates can be computed from the National Death Index instead of relying on hospitals to report mortality rates.<sup>27</sup> Models would need to be appropriately adjusted for case mix and there would be a lag in the ability to obtain these data, but we believe the advantages of using an independent data source outweigh its potential disadvantages.<sup>13-18,27</sup>

Where it is not possible to rely on independent data, it is important to minimize the role that subjective judgments can play in the data-generating process. The most controversial metric in the VHA access crisis required schedulers to determine and enter the date on which a patient wanted to be seen for follow-up care.<sup>28</sup> Though this process worked reasonably well most of the time, VHA audits found that some schedulers were instructed by supervisors not to enter the dates requested by patients.<sup>20</sup> An alternative way to measure waiting time is to calculate the duration between the date that the provider determines the patient should receive care and the date when he or she actually does. VHA physicians could enter these data directly into the electronic medical record and the data could be automatically extracted from the system. This process minimizes the role of discretion on the part of the scheduler.

The recommendation of the IOM to combine several measures into a composite metric is another important strategy to maintain data fidelity.<sup>2,3</sup> Composites can condense an overwhelming number of metrics while simultaneously making incentives less responsive to the manipulation of individual measures. For example, the VHA Office of Mental Health combines over 30 individual metrics into three composites to track mental health care performance across facilities and over

time.<sup>29</sup> In addition, the VHA clinical operations uses a scheduling composite metric that measures fidelity to scheduling policy.<sup>30</sup> Overall, the composite metrics are less easily manipulated than individual metrics but composites can hide data fidelity issues in specific individual metrics. This concern highlights the importance of validating all the individual metrics in the composite.

Finally, although careful validation and implementation can reduce the risk of unintended consequences, it is typically not possible to anticipate all strategic responses to new metrics and incentives. Consequently, ongoing monitoring is advisable to detect potential degradation of data fidelity over time. One way to accomplish this is to track multiple measures that have historically been highly correlated while incentives are linked to only a few. For example, the VHA is currently tracking the difference between two measures of waiting times for new patients, even though only one measure is publically reported.<sup>30</sup> Historically, these two measures have been closely related, so divergence between them either nationally or at a particular facility could indicate strategic changes in data generation processes.

### WHAT'S NEXT: DEVELOPING COMPREHENSIVE DATA TO FILL METRIC GAPS

The IOM report argues there is insufficient data to measure many of the core metrics. The identification of these gaps allows stakeholders to work on developing appropriate data to comprehensively measure outcomes.<sup>3</sup> The VHA is a leader in this regard with current plans to enhance its electronic medical record system. In concordance with the IOM's recommendation the new system is being designed with generation of performance metrics as a fundamental feature from the beginning instead of trying to compute metrics from whatever data were recorded after implementation. This approach will improve the collection of key data elements that measure health care value and increase the ease with which health record data can be shared between VHA and non-VHA systems.<sup>31</sup>

Access to non-VHA data is crucial to improving the comprehensiveness of outcome measures because veteran enrollees rely heavily on non-VHA care. In the 2013 Survey of Veteran Enrollees Health and Reliance Upon VA, only 30 % of respondents reported they relied solely on the VA to meet all their health care needs. Over half of the respondents reported they were also enrolled in Medicare, and almost 30 % reported they were enrolled in another individual or group health insurance plan.<sup>32</sup> Many outcomes, particularly acute ones such as heart attack and stroke, are more likely to be recorded in non-VHA records (because the nearest facility at the time of the event was not a VHA facility).<sup>33,34</sup> More comprehensive outcome measures better reflect the true health status of the patient population, which should be the focus of performance metrics.<sup>2</sup> Access to non-VHA data would also strengthen the

fidelity of performance measures by supplying more independently collected data. These data could be used directly in reported metrics or indirectly in alternative measures computed for fidelity monitoring purposes.

### CONCLUSION

The recent emphasis on value in health care has joined with advances in health data systems to fuel a rapid expansion in development and use of quantitative performance metrics. The IOM report correctly highlights the need to simplify the landscape of performance metrics by defining key health care goals through a consensus-building process and then obtaining needed data to measure these goals. In order to increase the chances that metrics will measure the agreed-upon goals, policymakers and managers should validate metrics before using them and employ implementation strategies that minimize unintended consequences. These strategies include maximizing the use of independent data sources, maintaining multiple related measures for monitoring purposes, and employing composite measures. Continuing evolution of electronic medical record systems should improve the design of performance metrics by providing more comprehensive data for outcome measurement.

---

**ACKNOWLEDGMENTS:** Funding for this research was provided by the VHA Access & Clinic Administration Program and CREATE 12-023 from the Health Services Research and Development Service, Department of Veterans Affairs. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs, Boston University, Harvard University or Northeastern University.

**Corresponding Author:** Julia C. Prentice, PhD; Health Care Financing and Economics VA Boston Healthcare System, 150 S. Huntington Avenue, Mailstop 152H, Boston, MA 02130, USA (e-mail: Julia.Prentice@va.gov).

**Compliance with Ethical Standards:**

**Conflict of Interest:** The authors report no conflicts of interest.

### REFERENCES

1. Lee TH. Putting the value framework to work. *N Engl J Med.* 2010;363:2481-3.
2. Blumenthal D, McGinnis J. Measuring vital signs: an IOM report on core metrics for health and health care progress. *JAMA.* 2015;313(19):1901-2.
3. IOM (Institute of Medicine). Vital signs: Core metrics for health and health care progress. Washington: The National Academies Press; 2015.
4. VA Needs Better Data on Extent and Causes of Waiting Times. GAO/HEHS-00-90. Washington DC: United States General Accounting Office; 2000.
5. Jaffe G, O'Keefe E. Obama accepts resignation of VA Secretary Shinseki. *Washington Post.* May 30, 2014.
6. Agency for Healthcare Research and Quality, Medical Expenditure Panel Survey available at <http://meps.ahrq.gov/mepsweb/>
7. VHA Office of Analytics and Business Intelligence, Patient Experience Reports and Scheduling available at <https://vaww.car.rtp.med.va.gov/programs/shep/sheReporting.aspx>
8. Prentice JC, Davies ML, Pizer SD. Which outpatient wait-time measures are related to patient satisfaction? *Am J Med Qual.* 2014;29(3):227-35.

9. 2009 Survey of Physician Appointment Wait Times, Physician Compensation, Salary and Physician Practice Surveys. Available at: <http://www.merrithawkins.com/compensation-surveys.aspx>. Accessed January 13, 2012.
10. **Resneck J Jr, Pletcher MJ, Lozano N.** Medicare, medicaid, and access to dermatologists: the effect of patient insurance on appointment access and wait times. *J Am Acad Dermatol.* 2004;50:85–92.
11. **Harris AH, Ellerbe L, Phelps TE, Finney JW, Bowe T, Gupta S, Asch SM, Humphreys K, Trafton J.** Examining the specification validity of the HEDIS quality measures for substance use disorders. *J Subst Abuse Treat.* 2015;53:16–21.
12. **Harris AH, Humphreys K, Bowe T, Tiet Q, Finney JW.** Does meeting the HEDIS substance abuse treatment engagement criterion predict patient outcomes? *J Behav Health Serv Res.* 2010;37(1):25–39.
13. **Prentice JC, Pizer SD.** Delayed access to health care and mortality. *Health Serv Res.* 2007;42(2):644–62.
14. **Prentice JC, Pizer SD.** Waiting times and ambulatory care sensitive hospitalization. *Health Serv Outcomes Res Methodol.* 2008;8(1).
15. **Pizer SD, Prentice JC.** What are the consequences of waiting for health care in the veteran population? *J Gen Intern Med.* 2011;26(Suppl 2):676–82.
16. **Prentice JC, Fincke BG, Miller DR, Pizer SD.** Primary care and health outcomes among older patients with diabetes. *Health Serv Res.* 2012;47(1 Pt 1):46–67 (**Published online 22 August 2011**).
17. **Prentice JC, Fincke BG, Miller DR, Pizer SD.** Outpatient waiting times and diabetes care quality improvement. *Am J Manag Care.* 2011;17(2):e43–54.
18. **Prentice JC, Davies ML, Pizer SD.** Using health outcomes to validate access quality measures. *Am J Manag Care.* 2013;19(11):e367–77.
19. **Caruso D.** Worst VA health care waits are in the South. Associated Press, April 9, 2015. <http://bigstory.ap.org/article/355ebf0f90fa4bbc9b58e1e5b9f8e3ec/va-makes-little-headway-fight-short-en-waits-care>
20. VA Access Audit & Wait Times Fact Sheet. System-Wide Overview. Washington, DC: US Department of Veterans Affairs, June 9, 2014.
21. Document: VA Audit Findings. CNN, June 9, 2014. Available at: <http://www.cnn.com/interactive/2014/05/politics/document-va-audit/>
22. Veterans Access, Choice and Accountability Act of 2014 (“Choice Act”). Fact Sheet. Washington DC: US Department of Veterans Affairs.
23. Veterans Access, Choice, and Accountability Act of 2014. HR 3230. Washington DC: US Congress, January 3, 2014.
24. VA Announces Single Regional Framework under MyVA Initiative. Washington, DC: US Department of Veterans Affairs, January 26, 2015.
25. **Kaplan R.** Robert McDonald Announces Plans to Reorganize the VA. CBS News. November 10, 2014.
26. **Roth S, Gonzales R, Harding-Anderer T, Bloom FJ, Graf T, Stahl MS, Maselli JH, Metlay JP.** Unintended consequences of a quality measure for acute bronchitis. *Am J Manag Care.* 2012;18(6):e217–24.
27. Center for Disease Control and Prevention. National Death Index available at <http://www.cdc.gov/nchs/ndi.htm>
28. Department of Veterans Affairs Access Audit. System-Wide Review of Access Results of Access Audit Conducted May 12, 2014 through June 3, 2014. Washington DC: US Department of Veterans Affairs.
29. Personal communication with Matthew Neuman, VHA Program Evaluation and Resource Center. December 1, 2014.
30. VHA Support Service Center (VSSC). Scheduling Trigger Tool. Available at: <http://reports2.vssc.med.va.gov/ReportServer/Pages/ReportViewer.aspx?%2fSysRD%2fWaitTimes%2fScheduling+Trigger+Tool&rs:Command=Render>
31. VistA Evolution — Initial Operating Capability (IOC) Fact Sheet. Washington DC: US Department of Veterans Affairs, December 1, 2014.
32. Assistant Deputy Undersecretary of Health. Survey of Veteran Enrollees’ Health and Reliance Upon VA. Available at <http://www.va.gov/healthpolicyplanning/analysis.asp>
33. **Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ.** Study outcomes and hospital utilization in the elderly: the advantages of a merged data base for medicare and veterans affairs hospitals. *Med Care.* 1992;30(5):377–91.
34. **Hendricks A, Gardner J, Frakt A, et al.** What can medicaid data add to research on VA patients? *J Rehabil Res Dev.* 2010;47(8):773–80.