Implementing Measurement Science for Electronic Health Record Use

Edward R. Melnick, MD, MHS  
Department of Emergency Medicine, Yale School of Medicine, New Haven, Connecticut, and Department of Biostatistics (Health Informatics), Yale School of Public Health, New Haven, Connecticut.

Christine A. Sinsky, MD  
American Medical Association, Chicago, Illinois.

Harlan M. Krumholz, MD, SM  
Center for Outcomes Research and Evaluation, Yale–New Haven Hospital, New Haven, Connecticut; and Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut.

Corresponding Author: Edward R. Melnick, MD, MHS, Department of Emergency Medicine, Yale School of Medicine, 464 Congress Ave, Ste 260, New Haven, CT 06519 (edward.melnick@yale.edu).

How did it happen that the electronic health record (EHR) emerged as a concern regarding clinicians’ well-being, and what is the path forward? Increasingly, evidence indicates that the EHR is imposing an intolerable burden on clinicians and may be degrading, rather than elevating, clinical care. A study of EHR use measurements across 2 vendor products found that ambulatory, nonteaching physicians (n = 573) spent more than 5 hours on the EHR for every 8 hours of scheduled clinical time.1 Clinicians trained in patient care are locked into hours of screen time to complete mandatory clinical and documentation tasks, often unrelated to the quality of the care.

In clinical quality, studies that revealed variation in practice and gaps in performance paved the way for programs to improve care and outcomes. Without the capacity to identify targets and evaluate interventions, quality was mired in an era of implicit review without systemic approaches to improvement. Measures are not sufficient for improvement, but good measures, in the proper context, have a central role in supporting and incentivizing better performance.

The EHR era is still in its infancy. An appreciation of its failures is widespread, but reliable, quantitative metrics of vendor performance are lacking. Scientifically sound EHR use metrics are needed to drive measurable improvements in EHR design, implementation, and regulation to inform improvements in clinical workflows and teamwork. For a measure to be scientifically sound, its results must be precise, reliable, valid, and adequately risk adjusted. Such measures can be used to compare vendors and instances of the same product, identify variation and best practices among clinicians, support training of students and residents, and spark efforts to improve. For example, EHR use metrics could determine (1) time spent on specific activities (eg, clinical documentation, inbox management, or work outside of scheduled clinical hours), (2) task switching between activities, (3) the number and rates of mouse clicks per task, and (4) the influence of teamwork on total clinician EHR time per day, to name a few. Currently, such tools are unable to assess what has become the most vital platform for the delivery of clinical care.

The nascent science of EHR use measurement has already started, albeit in a preliminary phase, and has evolved from collecting data via survey, self-timing, and direct observation to automated audit log capture at scale. Originally intended to track inappropriate access to protected health information, audit logs capture granular observational data (every screen viewed, keystroke, and mouse click) in an automated fashion. Interpreting log data, alone and in combination with other data sets, has the potential to show how safely, effectively, and efficiently the EHR is integrated into clinical care.

EHR audit logs are an appealing data source for measurement. Yet the validity and reliability of their data remain in question due to their unwieldy and subsequent inaccessible nature and lack of standard data definitions. Standard definitions of time-out lengths and work performed outside of scheduled clinical hours across vendor products and better integration of clinician schedules with EHR audit logs could begin to address many of these issues.

Despite current limitations, early EHR use measurement findings offer critical implications. For example, an unadjusted, longitudinal analysis in an academic health care system reported female ambulatory physicians across specialty disciplines (n = 389) spent 33 more minutes in the EHR per day than male physicians (n = 608).2 An analysis of 573 ambulatory physicians practicing in 2 health care systems using different vendor products confirmed this finding after normalizing EHR use to 8 hours of scheduled patient time and adjusting for physician age, specialty, health system, and number of hours scheduled per month.1 This difference in EHR use by physician sex suggests important areas for further inquiry and understanding.

The EHR has the potential for benefit, harm, and burden. To optimize EHR design, implementation, and regulation, EHR use measures must be developed that are trustworthy, clinically important, scientifically sound, transparent, and feasible for implementation. Improvements require reproducible measurements to track performance. The profession must be able to document EHR usability with valid and reliable assessments relevant to use in clinical settings. Results can engender accountability for how this technology is designed and implemented, identify specific opportunities for improvement, and support the evaluation of new strategies. For example, given the association between mortality and door-to-balloon time in ST-elevation myocardial infarction, in 2005 the Centers for Medicare & Medicaid Services (CMS) mandated reporting door-to-balloon time measurements for acute myocardial infarction. Over the next 5 years, door-to-balloon time declined from 96 to 64 minutes.3 Following a rigorous
development, validation, and implementation process, CMS's quality metrics allow accurate, reliable, and meaningful comparisons of care delivered within and across groups to assess, ensure, and improve care quality. Over the past 30 years, quality measurement and reporting have driven to enhance care consistency and transparency, with pipelines for accountability in performance.

There is no such accountability regarding EHR use. High switching costs and an absence of purchasing power for clinician users have created a misaligned market without the competition and market pressures that would ordinarily reward usability improvements. Existing regulatory oversight has failed to correct this and, in some circumstances, has contributed to the administrative burdens mediated through the EHR. Drawing attention to the EHR user experience with empirical measurements could add visibility and transparency to what is happening—the hours wasted and the risks incurred—that could stimulate usability improvements. Furthermore, the Office of the National Coordinator of Health IT could operationalize this by requiring regular reporting of scientifically sound EHR use measures by vendors to maintain their EHR's certification. The burden of reporting would be on vendors not clinicians.

Scientifically sound measurements could provide an accurate and meaningful characterization of the current state of EHR use, including comparisons within and across products, groups of clinicians, and individuals, as well as before and after policy, regulatory, or workflow changes. Benchmarking use in this way could guide evaluation and monitoring of data-driven practice and EHR redesign. For example, comparison between US and international users of the same EHR vendor product highlight major differences in total EHR time and time on specific tasks such as order entry and inbox that may reflect key differences in the regulatory and policy environments. Highlighting how much time is spent on documentation and other EHR tasks not directly related to clinical care could inform local redesign and future policy regarding the documentation, billing, or regulatory requirements.

Determining the ideal specifications for EHR use metrics grounded in measurement science will require a rigorous development and validation process analogous to the CMS and National Quality Forum's quality measure development process. These processes emphasize the creation of clinically important, scientifically sound that are feasible for implementation at a reasonable cost, effort, and time. For a measure to be clinically important, it must (1) have evidence linking it to a particular outcome, (2) target an area with poor performance and/or wide variation, and (3) produce actionable results that are usable and relevant to the needs and interests of specific stakeholders including clinicians, patients, practice leaders, vendors, and policy makers. Clinicians should find that EHR use measures reduce clerical burdens and are useful to improve their practice; practice leaders should find they improve care delivery and increase workforce retention; and vendors should find they drive product improvement and allow product comparison that can increase market share.

The EHR has the potential for benefit, harm, and burden. To optimize EHR design, implementation, and regulation, EHR use measures must be developed that are trustworthy, clinically important, scientifically sound, transparent, and feasible for implementation. These measures are needed now.