e-Measures: insight into the challenges and opportunities of automating publicly reported quality measures

Terhilda Garrido,1 Sudheen Kumar,1 John Lekas,1 Mark Lindberg,1 Dhanyaja Kadiyala,1 Alan Whippy,2 Barbara Crawford,3 Jed Weissberg4

ABSTRACT

Using electronic health records (EHR) to automate publicly reported quality measures is receiving increasing attention and is one of the promises of EHR implementation. Kaiser Permanente has fully or partly automated six of 13 the joint commission measure sets. We describe our experience with automation and the resulting time savings: a reduction by approximately 50% of abstractor time required for one measure set alone (surgical care improvement project). However, our experience illustrates the gap between the current and desired states of automated public quality reporting, which has important implications for measure developers, accrediting entities, EHR vendors, public/private payers, and government.

INTRODUCTION

Quality measurement, a key lever to improve healthcare, has traditionally relied on administrative claims data and time-consuming manual chart abstraction.1 Health information technology (IT) promises to generate quality measurement and public reporting through automated data collection more readily.5

Many believe that electronic health records (EHR) offer new potential for quality measurement.1 The US Health Information Technology for Economic and Clinical Health Act of 2009 invested US$20 billion for health IT infrastructure and Medicare and Medicaid ‘meaningful use’ (MU) incentives. The proportion of acute care hospitals adopting at least a basic EHR more than doubled between 2009 and 2011, and in 2011 85% of hospitals planned to attest to MU of certified EHR technology by 2015, which includes submitting clinical quality measures.4 Stage 1 measures include 15 metrics focusing on emergency department (ED) throughput, stroke care, venous thromboembolism (VTE) prevention, and anticoagulation.5

The taxonomy for measuring and reporting performance from EHR is evolving; we use ‘e-measures’ and ‘automated measures’ interchangeably to refer to all partly or fully automated processes for generating performance information from EHR-contained data. Multiple stakeholders are vested in their development: the consumers and communities e-measures are intended to benefit, governmental entities, and commercial payers.1 At the request of the US Department of Health and Human Services, the National Quality Forum convened diverse stakeholders to provide recommendations for retooling 113 paper-based measures to an electronic format.6 e-Measure implementation occurs at the interface between measure developers and providers using EHR. Understanding the work of retooling paper-based quality measures for automated reporting illuminates the gap between the current and desired states of e-measures; we report here Kaiser Permanente’s experience with automating quality reporting.

AUTOMATED QUALITY REPORTING AT KAISER PERMANENTE

Kaiser Permanente’s EHR, KPHealthConnect, enables clinicians and employees to manage the healthcare and administrative needs of nine million members across eight geographic regions in a seamless and integrated way, with resulting quality and efficiency benefits.2,7 Implemented beginning in 2003, KP HealthConnect is deployed program-wide; individual regions customize EHR builds to local conditions and needs.

In 2010, Kaiser Permanente care reporting staff began to re-tool the joint commission (TJC) core measures for automated quality reporting. The Northern California, Southern California and Northwest regions currently use or are developing e-measures. Each measure comprises numerous data elements identifying population inclusion and exclusion and outcomes criteria.

Kaiser Permanente selected measures for automation on the basis of their clinical significance, importance to regulatory reporting, and reliance on discrete data elements: unambiguous numeric or coded values, such as cardiac rate and rhythm. Discrete data greatly reduce (but do not eliminate) the possibility of inaccurate reporting. They enable semantic understanding of the nature of the underlying information. In contrast, non-discrete data fields contain strings of indistinguishable characters; an example is the VTE-5 measure: VTE warfarin therapy discharge instructions. Documentation takes the form of embedded text, and text mining or current natural language processing cannot meet the current 100% accuracy requirement for publicly reported measures.

To date, Kaiser Permanente has focused on six core measure sets: acute myocardial infarction, ED patient flow, immunizations, the surgical care improvement project, pneumonia, and VTE prophylaxis. Kaiser Permanente has partly or fully automated 21 of 29 measures in these sets. The number of data elements per measure ranges from eight to 93, and the proportion that is discrete and mapped ranges from 43% to 100% (table 1). We
derive the balance of TJC core measure reports from administrative claims data and manual chart abstraction.

Figure 1 represents the iterative five-step process of developing and maintaining selected measures for automated reporting. A more extended discussion is available online (see supplementary web-only appendix 1, available online only).

Core measure interpretation

Measures first require interpretation in light of current specifications. Domain experts, such as abstractors and analysts, informaticists, legal and compliance staff, and clinical experts, provide input. Measures interpretation must be consistent throughout Kaiser Permanente.

Mapping

Mapping links specifications of the interpreted measure to EHR data tables. EHR database design is highly complex and its relationship to clinical documentation is often obscure. In addition, regional configurations and local documentation workflows vary.

Coding

Coding extracts required data from the mapped data tables. Convergent medical terminology ‘groupers’—lists of similar medications maintained over time—are intended to eliminate the need to ‘hard-code’ the constantly evolving universe of individual medications.

Quality assurance/validation

Validation of automated quality reports ensures accuracy and is conducted by comparing automated results to the official submission and re-examining interpretation, mapping, and coding to rectify discrepancies.

Maintaining code over time

TJC measure definitions change to reflect evolving evidence and clinical practice, and EHR updates and new releases may change database design. Diagnostic nomenclature, procedural codes, and medication identifiers may change, as can internal factors impacting automated quality reporting.

RESULTS OF AUTOMATED QUALITY REPORTING

With validated accuracy of 100%, potential gains to the organization result from increased efficiency. To assess these, we compared pre and post-automation extraction time.

We measured the time required to abstract each of 20 randomly selected cases using manual abstraction methods. After partial or full automation was complete, we again measured the time required to abstract each case. Our calculations omit the time required for automation; for example, an 8-week collaborative effort between national reporting and regional abstraction staff fully automated five individual immunizations measures.

Table 1 contains the average time savings per case from using partly or fully automated reporting, compared to manual chart abstraction. In addition, we estimated the costs of development and ongoing maintenance for these measures, which included time for programmers, registered health information technicians (chart abstractors), and management and oversight, as well as a portion of IT infrastructure (eg, hardware and software). We found a breakeven in savings over cost by the fourth year, by then achieving an ongoing savings stream of approximately US $1 million annually for one Kaiser region.

DISCUSSION

Our experience illustrates opportunities and challenges inherent in e-meaures and gaps between the current and potential states of automated quality reporting. The five-step process we describe can serve as a template for the development of automated measures elsewhere, but it is unlikely to expedite the local development process. The time savings we observed highlight substantial opportunities for increased efficiency to augment the intrinsic benefits of performance reporting. Overhead costs of public quality reporting are significant. For example, Kaiser Permanente has reported on 50 well-established metrics for 15 years; the annual cost is approximately US$6.75 million, excluding expenses related to IT systems, storage, and oversight (unpublished data, Kaiser Permanente, 2010). We observed an approximate 50% reduction in abstraction time just for the partly automated surgical care improvement project measures; this time savings is likely to be broadly achievable.

Partial abstraction saves time over a completely manual process, expanding the capacity of existing abstraction staff and allowing us to forego hiring additional abstractors despite an expanding number of quality measures. Each automated element has been rigorously tested and validated by subject matter experts and needs no manual review for confirmation. In addition, even in some instances when we cannot fully automate a field, we supply a ‘trigger location’ within the EHR for confirmation by manual review, changing abstraction from an intuitive search to a focused verification.

| Table 1 Current state of automated quality reporting at Kaiser Permanente Northern California |
|-------------------------------|-----------------|-----------------|-----------------|-------------------|
| TJC core measure               | Required data fields | Mapped discrete data fields | Mapped data elements (%) | Time saved per case in minutes |
| SCIP                          | 93               | 43              | 46              | 11                |
| AMI                           | 41               | 20              | 49              | 10                |
| PN                            | 41               | 28              | 68              | 5                 |
| IMM                           | 15               | 15              | 100             | 5                 |
| VTE                           | 67               | 23              | 43              | 14                |
| ED                            | 8                | 8               | 100             | 5                 |

AMI, acute myocardial infarction; ED, emergency department; IMM, immunizations; PN, pneumonia; SCIP, surgical care improvement project; TJC, the joint commission; VTE, venous thromboembolism.
The number of contemplated and required public reporting initiatives is growing exponentially. Automated quality reporting offers an opportunity to obtain the transparency and accountability benefits of increased reporting without adding to high overhead costs or depleting organizational quality improvement budgets. In addition, when quality measures are automated, some data are available as real-time organizational intelligence, expediting care improvement cycles.

Automated reporting can also be extended to quality measures that are not uniformly publicly reported; for example, surgical site infections (SSI). The Centers for Medicare and Medicaid Services requires hospitals to report infections related to colon surgeries and abdominal hysterectomies to the National Healthcare Safety Network for eventual publication on the Centers for Medicare and Medicaid Services Hospital Compare website. Although state policies vary, the California Department of Public Health requires SSI reporting for 29 National Healthcare Safety Network-defined procedures, with data published online. An automated SSI reporting process reduced Kaiser Permanente’s manual surveillance full time equivalent by 80%, reflecting an aggregate savings of US$2 million (see supplementary appendix 2, available online only).

However, our experience also highlights the challenges of automating quality reporting. A recent report questions the accuracy of automated reporting across conditions. Achieving 100% accuracy required additional staff time to rectify discrepancies between methods; this time would probably be required in all settings.

Although little has been reported on automated quality reporting across conditions, our experience is supported by existing evidence. A primary challenge is that EHR were not initially designed to calculate, compile, and report on quality measures. Their core function is to capture, store, and track clinical data to support transactions. Consequently, much of the quality reporting supported by EHR is not fully automated. EHR meeting explicit stage 1 MU data capture requirements are estimated to provide approximately 35% of needed data. EHR also including electronic physician notes and medication administration records may provide up to 65% of needed data. Similarly, across the measures reported here, an average of 61% of needed data was available as discrete elements. Our experience is likely to be generalizable to a large extent, as we rely on data that are available to providers during the course of an encounter. All certified EHRs are likely to have these data. Some types of data lend themselves to discrete representation: vital signs, medications, and coded diagnoses and procedures, for instance. Other types of data are intrinsically more nuanced, such as discharge teaching. Our priority is to capture data that are a natural ‘exhaust’ from provider workflows, rather than asking physicians and nurses to interrupt natural care processes to complete a template or form.

Solutions to bridge the gap between the current and desired states of automated quality reporting are complex. One option, full automation, requires that all data elements are represented by standardized terminologies and codes within an EHR system and that the same standards are used locally and nationally. This is impractical. Many data elements that are difficult or impossible to automate are also essential for measure meaningfulness. For instance, an acute myocardial infarction measure relates to smoking cessation counseling, which is recorded narratively in progress notes or a teaching summary, precluding automation. If measures lack sufficient meaningfulness, physicians and other clinicians will have less incentive to drive operational and practice changes to improve performance on them.

Strong federal involvement and guidance would be required to achieve a highly coordinated approach to addressing gaps in automated quality measurement standards and processes; this may risk generating reporting requirements that inappropriately drive cumbersome clinical workflows. A collaborative approach is evidenced by numerous stakeholders at the national level, including the Agency for Healthcare Quality and Research and, as noted earlier, the National Quality Forum, working together and independently to advance e-measures. However, the continuing development of quality measures should not embrace automation at the cost of meaningful clinical detail or overburdening clinician workflow.

Many questions remain. Even with improved standardization of terminologies and codes, EHR content, structure, and data format vary, as do local data capture and extraction procedures. Within a single institution, significant differences in denominators, numerators, and rates arise from different electronic data sources, and documentation habits of providers vary. Data entered into the EHR may not be interpreted or recognized, resulting in substantial numerator loss and underestimates of the delivery of clinical preventive services. EHR vendors can potentially support e-measures. However, organizations typically customize vendor-provided builds and workflows are local; data extraction will require local customization, too. Structured clinical data are often captured electronically through clinical reminders that are relatively insensitive to context and interfere with workflows, and adapting workflows to support documentation is short sighted. For instance, data supporting the medication reconciliation measure for stage 2 MU can be found in medication actions, such as new, changed, discontinued, or adjusted medications or a refill request, from which review and reconciliation can be inferred. A vendor-generated solution might be a check box for ‘medication reconciliation complete.’ This both adds an inefficient step to the workflow and creates the possibility of providers indicating that reconciliation occurred without conducting it.

A final comment pertains to the difference between retooling quality measures that were designed for manual abstraction and developing quality measures for internal use de novo, with which Kaiser Permanente has robust experience. Given the level of exactitude required to retool manual measures for automation, de-novo development of quality measures can potentially be more straightforward. However, our experience is that the latter also incurs substantial development costs to capture variability in clinical workflows.

CONCLUSIONS
Kaiser Permanente has fully or partly automated six TJC measures. Time savings from automation were substantial, but our experience illustrates the complex nature of this undertaking and the gap between the current and desired states of automated quality reporting. The goal of fully automating quality measurement may challenge the goals of supporting provider-driven efficient workflows and retaining the meaningfulness of quality measures.

Correction notice This paper has been made unlocked since it was published Online First.

Acknowledgements The authors acknowledge the support and leadership of the following members of the Kaiser Permanente quality and informatics teams: Greg Huffman, current director of care reporting, Randy Watson, quality data manager; Lorelle Poropat, practice leader for regional data consulting; and Brian Hoberman, MD, physician leader for Northern California KP HealthConnect. In addition, dozens of experts and providers help support the ongoing effort to automate publicly...
reported quality measures at regional and national levels, including Andy Amster, director of the Center for Health Care Analytics.

Contributors TG, JW, AW and BC sponsored the work. TG and AW initiated the work. TG, SK, AW and BC designed the structure of the work and plan. SK, ML, DK and JL did the programing, automation, and quality assurance. TG, AW and BC removed barriers as the work progressed. TG and SK designed the data collection. SK, ML, DK and JL conducted the data collection. TG, SK, ML, DK and JL documented the transformation process. TG and JW drafted the initial paper. TG, JW, AW, BC, SK and DK reviewed and provided comments. TG is the guarantor.

Contributors TG, JW, AW and BC sponsored the work. TG and AW initiated the work. TG, SK, AW and BC designed the structure of the work and plan. SK, ML, DK and JL did the programing, automation, and quality assurance. TG, AW and BC removed barriers as the work progressed. TG and SK designed the data collection. SK, ML, DK and JL conducted the data collection. TG, SK, ML, DK and JL documented the transformation process. TG and JW drafted the initial paper. TG, JW, AW, BC, SK and DK reviewed and provided comments. TG is the guarantor.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES


