ADVANCING CLINICAL DECISION SUPPORT (CDS) AND ELECTRONIC CLINICAL QUALITY MEASUREMENT (ECQM)

by

Polina Kukhareva

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biomedical Informatics

The University of Utah

December 2017

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Polina Kukhareva**

has been approved by the following supervisory committee members:

Kensaku Kawamoto	, Chair	**09/05/2017** Date Approved
Catherine J. Staes	, Member	**09/05/2017** Date Approved
Charlene R. Weir	, Member	**09/05/2017** Date Approved
Stanley M. Huff	, Member	**09/05/2017** Date Approved
Howard R. Weeks	, Member	**09/05/2017** Date Approved

and by **Wendy W. Chapman** , Chair/Dean of

the Department/College/School of **Biomedical Informatics**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Clinical decision support (CDS) and electronic clinical quality measurement (eCQM) are 2 important computerized strategies aimed at improving the quality of healthcare. Unfortunately, computer-facilitated quality improvement faces many barriers. One problem area is the lack of integration of CDS and eCQM, which leads to duplicative efforts, inefficiencies, misalignment of CDS and eCQM implementations, and lack of appropriate automated feedback on clinicians' performance. Another obstacle in the acceptance of electronic interventions can be the inadequate accuracy of electronic phenotyping, which leads to alert fatigue and clinicians' mistrust of eCQM results.

To address these 2 problems, the research pursued 3 primary aims:

Aim 1. Explore beliefs and perceptions regarding the integration of CDS and eCQM functionality and activities within a healthcare organization.

Aim 2. Evaluate and demonstrate feasibility of implementing quality measures using a CDS infrastructure.

Aim 3. Assess and improve strategies for human validation of electronic phenotype evaluation results.

To address Aim 1, a qualitative study based on interviews with domain experts was performed. Through semistructured in-depth and critical incident interviews, stakeholders' insights about CDS and eCQM integration were obtained. The experts

identified multiple barriers to the integration of CDS and eCQM and offered advice for addressing those barriers, which the research team synthesized into 10 recommendations.

To address Aim 2, the feasibility of using a standards-based CDS framework aligned with anticipated electronic health record (EHR) certification criteria to implement electronic quality measurement (QM) was evaluated. The CDS-QM framework was used to automate a complex national quality measure at an academic healthcare system which had previously relied on time-consuming manual chart abstractions.

To address Aim 3, a randomized controlled study was conducted to evaluate whether electronic phenotyping results should be used to support manual chart review during single-reviewer electronic phenotyping validation. The accuracy, duration, and cost of manual chart review were evaluated with and without the availability of electronic phenotyping results, including relevant patient-specific details. Providing electronic phenotyping results was associated with improved overall accuracy of manual chart review and decreased review duration per test case.

Overall, the findings informed new strategies for enhancing efficiency and accuracy of computer-facilitated quality improvement.

TABLE OF CONTENTS

LIST OF TABLES

Tables

LIST OF FIGURES

Figures

ACKNOWLEDGMENTS

Firstly, I would like to express my genuine gratitude to my scientific advisor Dr. Kensaku Kawamoto for the constant support of my PhD study and related research, for his patience, inspiration, and enormous knowledge. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my PhD study.

I would also like to particularly thank Dr. Catherine Staes for support, advice and ideas which helped me with this dissertation. I would like to thank the rest of my dissertation committee: Dr. Charlene Weir, Dr. Stanley M. Huff, and Dr. Howard R. Weeks for their insightful comments and encouragement.

I thank my fellow colleagues from the Knowledge Management and Mobilization Team for the useful advice, shared work experience, and for all the fun we have had in the last four and a half years.

Additionally, I would like to thank all the professors, postdocs and students at the University of Utah Department of Biomedical Informatics for creating a healthy and supportive environment.

Finally yet importantly, I would like to thank my family: my husband, my parents and my sister for supporting me spiritually throughout writing this dissertation.

CHAPTER 1

INTRODUCTION

1.1 Clinical Quality Improvement Strategies

Delivering quality healthcare is challenging due to ongoing and ubiquitous variation in health system processes that may lead to errors.[1] Measuring and reducing variation from evidence-based clinical recommendations have been shown to improve quality and decrease costs of healthcare.[2] The increasing adoption of electronic health records (EHRs) and associated interoperability standards in recent years has created a foundation upon which structured electronic data can be used to facilitate quality improvement strategies such as CDS and clinical quality measurement (CQM).

1.1.1 Clinical Decision Support (CDS)

A substantial body of evidence shows that, if correctly implemented, CDS could be effective in improving clinical and process outcomes.[3] Initially, many large academic hospitals developed their own EHRs and their own CDS Systems (CDSSs). Later, when home-grown EHR systems were replaced by commercial EHR systems, those CDSSs could not be easily adopted because they were tightly coupled with the home-grown EHR systems for which they were developed. Currently, many CDSSs are built on top of the

customized implementations of commercial EHRs specific to a given healthcare organization.

Kawamoto et al. have previously suggested that a standards-based, service-oriented architecture could be used to make CDS logic sharable between different EHRs.[4] In pursuing this potential approach to CDS, a promising resource is an open-source, standards-based, service-oriented framework for CDS known as OpenCDS.[5] An EHR system can submit patient data to OpenCDS and obtain patient-specific assessments and recommendations that are provided to clinicians via alerts, reminders, or other CDS modalities.[6] OpenCDS is compliant with the HL7 Virtual Medical Record (vMR) and HL7 Decision Support Service (DSS) standards, and it leverages various open-source component resources, including the JBoss Drools knowledge management platform and Apelon Distributed Terminology System.

1.1.2 Electronic Clinical Quality Measurement (eCQM)

Clinical quality measures are measures of processes, experiences, and/or outcomes of patient care. Having a means to assess healthcare quality is essential for identifying deviations from evidence-based best practices and mitigating preventable errors.[7] Currently, CQM is required by public and private payers, regulators, accreditors and others that certify performance levels for consumers, patients and payers.[8]

Current quality measurement systems in many hospitals include time-consuming manual paper and electronic record abstraction by a quality improvement specialist.[9,10] At large academic medical centers such as University of Utah Health Care (UUHC), manual data abstraction is often followed by data analysis performed by an external organization

such as the University HealthSystem Consortium.[11,12] There are several limitations with this process. For example, (a) 3 to 6 months may elapse between the time of a clinical procedure (eg, a surgery) and the time when feedback is given to a clinician; (b) human errors may be introduced during manual record review; and (c) only a subset of the patients and clinical events is oftentimes selected for review, leading to gaps in quality assessment coverage. Theoretically, the above problems could be solved using electronic clinical quality measurement (eCQM).

There are increasing mandates and financial incentives to use EHRs to measure quality as opposed to employing traditional manual processes for QM.[7,13] For example, the Meaningful Use (MU) recommendations issued by the federal Health Information Technology Policy Committee in November 2012 require the implementation of eCQM as well as CDS for high-priority conditions, and the use of related standards.[13,14] One of the promises of implementing EHRs is the possibility for automatic generation of eCQM.[10] A MU-certified EHR must be able to export standardized quality reports, which can then "be fed into a calculation engine to compute various aggregate scores."[15] Following these recommendations, major EHR vendors such as Epic have started to integrate eCQM logic into their products.[16]

Currently, however, only some EHR vendors offer quality measurements embedded in their system, and the scope of measures supported is not always comprehensive.[10,16] For example, a study in 2010 found that KPHealth Connect had automated 6 of 13 Joint Commission measurement sets.[10] As well, EpicCare Inpatient 2014 and EpicCare Ambulatory 2014 offered 56 National Quality Forum (NQF) quality measures required for MU certification out of over 700 NQF-endorsed measures on their

website.[16] While vendor-based solutions may be comprehensive in the scope of patients analyzed, their implementation may be a "black box" where the inner workings of the algorithms employed are difficult to discern. Also, it is not always clear which version of each rule has been implemented or whether the quality measure logic is up-to-date. In addition, users may not have control over the logic to customize quality measurement. Even so, automated eCQM has the potential to provide quality reports on demand, may avoid human errors in manual abstraction, and can analyze *100%* of relevant patients and their encounters, as opposed to analyzing only a *subset* when using manual phenotyping. Most ongoing efforts to produce automated quality measures are tied to a specific EHR system, and the executable logic for the quality measure is not sharable between different EHR systems.[10]

1.2 Challenges Facing Quality Improvement Efforts

Despite multiple efforts undertaken to improve healthcare quality since the publication of the Institute of Medicine reports "To Err Is Human"[17] and "Crossing the Quality Chasm"[18], the quality of healthcare in the United States continues to be compromised by unnecessary variation in the implementation of clinical practice guidelines. Deficiencies in CDS and eCQM design, implementation, and maintenance, as well as misaligned incentives, can cause the low effectiveness of CDS and eCQM. For example, a meta-analysis of 26 papers showed that usability flaws in medication alerting systems have negative impact on workflow, technology effectiveness, medication management processes, and patient safety.[19] Informatics-based quality improvement efforts often fail to reach their goal because of multiple issues as summarized in Table

1.1. This dissertation research addressed 2 of these challenges: the lack of integration between CDS and eCQM and the low accuracy of phenotyping.

CDS and eCQM were traditionally implemented in silos and discussed separately in the medical literature. Only 3 out of 160 randomized clinical trials described in a systematic review by Lobach et al. describe CDSSs accompanied by periodic performance feedback, possibly because feedback requires additional development effort and could not be easily integrated with CDS.[3]

Once implemented, both CDS and eCQM need to be regularly reviewed and potentially updated. When they are programmed separately, maintenance of the logic requires duplication of effort. In addition, CDS and eCQM logic may get updated asynchronously or differently, which could cause confusion among clinicians. These issues may be exacerbated by differences in the background of personnel performing quality oversight compared to the technical personnel tasked with implementing decision support. Integration may be difficult when the incentives and mission are misaligned between the 2 teams.

Furthermore, validation processes for both CDS and eCQM need to be improved. Studies have shown that electronically reported MU quality measures have low accuracy.[20] Similarly, studies have shown that CDS use is compromised by alert fatigue and low attention of clinicians to some CDS alerts, partly due to poor accuracy of alerts.[19]

1.3 Potential Solutions

Potential solutions have been mapped to the challenges described above (Table 1.1).

1.3.1 Integration of CDS and eCQM

CDS and eCQM are highly related, as eCQM focuses on who is eligible for a needed intervention (denominator identification) and who among them has received the needed intervention (numerator identification), whereas CDS focuses on who is eligible for a needed intervention and has not received the needed intervention (equivalent to numerator identification). However, to the best of our knowledge, there have been limited reports of evaluation and validation in the literature concerning how technical approaches for one problem space can be reused in the other, especially pertaining to standards-based approaches. This finding is important because EHR certification criteria will likely require more automation and need for validation in the future.

It has been previously suggested that CDS and eCQM could be combined.[21] Furthermore, there has been a trend towards viewing CDS and eCQM as two sides of the same coin. There was a qualitative field study performed at the Regenstrief Institute, Partners Health Care System, and Veterans Health Administration that showed a paradigm shift from viewing CDS and performance measures as 2 separate approaches to viewing a clinical reminder as a real time performance measure with an "n of one."[22]

It has been shown that clinical reminders corresponding to performance measures could improve organizational performance.[3] Diabetes care was shown to improve significantly when a multifaceted intervention combining reminders and performance feedback was introduced.[23] This finding is congruent with the findings from a systematic review by Forrest et al. In this systematic review of randomized controlled trials for patients with type 2 diabetes, Forrest et al. found that CDS only improves patient outcomes when combined with feedback on performance.[24]

New methods are currently being developed to unify CDS and eCQM and follow the success of clinical pathways implementation.[25] There have been efforts to combine CDS and eCQM logic. For example, one of the proposed solutions is to use the National Quality Forum (NQF) Quality Data Model (QDM) and JBoss Drools rules engine.[26–28] However, these efforts are often not standards-based, and no conceptual framework was developed.[29–31] We hypothesized that technical integration of CDS and eCQM could be reached by leveraging a standards-based CDS Web service across a population for both eCQM and CDS.

In pursuing the integration of CDS and eCQM, it is important to understand the viewpoint and experience of different stakeholders, such as members of institutional quality teams and CDS teams. Thus, we proposed to investigate both cultural and technical challenges preventing CDS and eCQM integration and to develop a framework which would allow implementing CDS and eCQM simultaneously.

1.3.2 Improving Strategies for Validating Results of Electronic Phenotyping

Computable phenotyping entails automatic identification of patient records satisfying specific conditions. Computable phenotyping is essential for CQM, CDS, risk adjustment, clinical registries, predictive analytics, public reporting, and cohort identification for clinical trials and research.[32] Accuracy of such phenotyping is essential for CDS and eCQM to be optimally effective. For example, a time-series analysis at a large internal medicine practice using a commercial EHR showed that making point-of care reminders and feedback more accurate accelerated the rate of quality improvement.[33]

The testing of electronic phenotyping algorithms is important to detect errors and

provide high quality results over time. Double human chart review is often considered a "gold standard" of phenotyping validation in research and academic settings[34–37]; however, it is too expensive and slow to be used in operational settings.[38] Human review is subject to error and produces both false negative and false positive results when used to detect errors. This dissertation aims to develop a single human review-based phenotyping validation approach that is both pragmatic and high-performing.

Currently, there is no standard framework for electronic phenotyping validation. Newton et al. presented recommendations for phenotyping algorithms validation but did not focus on human expert review.[39] While validating quality measures for enterprise implementation at UUHC, our group initially developed an ad hoc validation methodology that was not sufficiently robust. We neither selected cases randomly, nor did we ensure an adequate mix of positive and negative results. To improve the quality of our validation strategy, we developed and formally evaluated a new and more robust electronic phenotyping validation framework.

1.4 Dissertation Aims

To address the problems raised above, the research had 3 primary aims:

Aim 1. Explore beliefs and perceptions regarding the integration of CDS and eCQM functionality and activities within a healthcare organization.

Aim 2. Evaluate and demonstrate feasibility of implementing quality measures using a CDS infrastructure.

Aim 3. Assess and improve strategies for human validation of electronic phenotype evaluation results.

Table 1.1 Challenges facing electronic quality improvement efforts and potential solutions

Challenge	Description	Potential Solutions
Poor user interface design	Unclear text, too many clicks to access the information, requested actions do not correspond to what the user requested.	Conduct usability testing.
Lack of interoperability[40]	Most existing CDS and eCQM systems and their knowledge bases have limited portability.	Employ standard-based approaches.
Lack of technical approaches to co-implement CDS and eCQM	Alerts are often not updated properly. The lack of standardization and poor versioning causes divergent CDS and eCQM implementations. Clinicians do not get feedback on their decisions.	Develop and evaluate technical approaches for integrating CDS and eCQM.
Lack of organizational integration between CDS and eCQM teams	Quality teams include analysts with a mission to evaluate and improve care quality. CDS teams include technical implementers with a mission to develop and implement functionality.	Pursue efficient integration of quality and CDS teams.
Clinicians do not see a need for computerized quality improvement.[41]	Clinician self-assessment of delivered care quality is often higher than their true performance.	Provide feedback on performance. It has been shown that feedback on performance lowers canceling of alerts by junior-level physicians.[42]
Low accuracy of electronic phenotyping	High number of false positive results causes alert fatigue and mistrust of quality measures.	Improve validation strategies.
Misaligned incentives[43]	Fee-for-service reimbursement models are still the predominant form of US healthcare reimbursement.	Align financial incentives with quality and outcomes (eg, via pay-for-performance).
Outcomes of interventions often not monitored or evaluated	Changes caused by quality improvement interventions are often not analyzed properly, thereby limiting opportunities for learning and continuous improvement.	Improve outcome evaluation.
Poor timing (reactive versus proactive)	CDS often appears after the user has already made a decision. Feedback from the QM can also be delayed and may be delivered months after the fact.	Improve the timing for presenting feedback within the user's workflow.
High cost	Implementing CDS and eCQM capabilities is oftentimes difficult and costly, with the need for highly skilled personnel.	Increase interoperability and collaboration to efficiently share CDS and eCQM capabilities (eg, both within and across institutions).

1.5 References

1. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med. 2010;363(22):2124-2134. doi:10.1056/NEJMsa1004404.

2. James BC, Savitz LA. How Intermountain trimmed health care costs through robust quality improvement efforts. Health Aff (Millwood). 2011;30(6):1185-1191. doi:10.1377/hlthaff.2011.0358.

3. Lobach D, Sanders GD, Bright TJ, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. Evid Rep Technol Assess (Full Rep). 2012;(203):1-784.

4. Kawamoto K, Jacobs J, Welch BM, et al. Clinical information system services and capabilities desired for scalable, standards-based, service-oriented decision support: consensus assessment of the Health Level 7 clinical decision support Work Group. AMIA Annu Symp Proc. 2012;2012:446-455.

5. OpenCDS Home. http://www.opencds.org/. Accessed January 30, 2014.

6. Kawamoto K, Shields D, Del Fiol G. OpenCDS: enabling clinical decision support at scale through open-source, standards-based software and resources. AMIA poster. 2011:1830.

7. Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. N Engl J Med. 2010;363(6):501-504. doi:10.1056/NEJMp1006114.

8. Meyer GS, Nelson EC, Pryor DB, et al. More quality measures versus measuring what matters: a call for balance and parsimony. BMJ Qual Saf. 2012;21(11):964-968. doi:10.1136/bmjqs-2012-001081.

9. Silow-Carroll S, Edwards JN, Rodin D. Using electronic health records to improve quality and efficiency: the experiences of leading hospitals. Issue brief (Commonwealth Fund).

10. Garrido T, Kumar S, Lekas J, et al. e-Measures: insight into the challenges and opportunities of automating publicly reported quality measures. J Am Med Inform Assoc. 2014;21(1):181-184. doi:10.1136/amiajnl-2013-001789.

11. University HealthSystem Consortium (UHC). https://www.uhc.edu/. Accessed January 30, 2014.

12. University of Utah Health Care - Salt Lake City, Utah. http://healthcare.utah.edu/about/. Accessed January 30, 2014.

13. Blumenthal D. Launching HITECH. N Engl J Med. 2010;362(5):382-385. doi:10.1056/NEJMp0912825.

14. Joseph AM. The stage 3 meaningful use preliminary recommendations: concerns are being raised. MLO Med Lab Obs. 2013;45(7):64.

15. Dolin RH. Accuracy of electronically reported "meaningful use" clinical quality measures. Ann Intern Med. 2013;159(1):73. doi:10.7326/0003-4819-159-1-201307020-00017.

16. Epic: Meaningful Use Stage 2 Certification Details. https://www.epic.com/software-certification.php. Accessed February 3, 2014.

17. National Research Council. To Err Is Human: Building A Safer Health System - Institute of Medicine. Washington, DC: The National Academies Press; 2000. http://www.iom.edu/Reports/1999/to-err-is-human-building-a-safer-health-system.aspx. Accessed February 26, 2014.

18. National Research Council, Committee on Quality of Healthcare in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press, D.C.: NATIONAL ACADEMY PRESS; 2001. http://www.nap.edu/catalog.php?record_id=10027. Accessed February 26, 2014.

19. Marcilly R, Ammenwerth E, Roehrer E, Pelayo S, Vasseur F, Beuscart-Zéphir M-C. Usability flaws in medication alerting systems: impact on usage and work system. Yearb Med Inform. 2015;10(1):55-67. doi:10.15265/IY-2015-006.

20. Kern LM, Malhotra S, Barrón Y, et al. Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study. Ann Intern Med. 2013;158(2):77-83. doi:10.7326/0003-4819-158-2-201301150-00001.

21. Lobach DF, Kawamoto K, Anstrom KJ, et al. Proactive population health management in the context of a regional health information exchange using standards-based decision support. AMIA Annu Symp Proc. January 2007:473-477.

22. Haggstrom DA, Saleem JJ, Militello LG, Arbuckle N, Flanagan M, Doebbeling BN. Examining the relationship between clinical decision support and performance measurement. AMIA Annu Symp Proc. 2009;2009:223-227.

23. Weber V, Bloom F, Pierdon S, Wood C. Employing the electronic health record to improve diabetes care: a multifaceted intervention in an integrated delivery system. J Gen Intern Med. 2008;23(4):379-382. doi:10.1007/s11606-007-0439-2.

24. Cleveringa FGW, Gorter KJ, van den Donk M, van Gijsel J, Rutten GEHM. Computerized decision support systems in primary care for type 2 diabetes patients only improve patients' outcomes when combined with feedback on performance and case management: a systematic review. Diabetes Technol Ther. 2013;15(2):180-192. doi:10.1089/dia.2012.0201.

25. Konrad R, Tulu B, Lawley M. Monitoring adherence to evidence-based practices:

a method to utilize HL7 messages from hospital information systems. Appl Clin Inform. 2013;4(1):126-143. doi:10.4338/ACI-2012-06-RA-0026.

26. Pathak J, Bailey KR, Beebe CE, et al. Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium. J Am Med Inform Assoc. 2013;20(e2):e341-8. doi:10.1136/amiajnl-2013-001939.

27. Li D, Endle CM, Murthy S, et al. Modeling and executing electronic health records driven phenotyping algorithms using the NQF Quality Data Model and JBoss® Drools Engine. AMIA Annu Symp Proc. 2012;2012:532-541.

28. Thompson WK, Rasmussen L, Pacheco JA, et al. An evaluation of the NQF quality data model for representing electronic health record driven phenotyping algorithms. AMIA Annu Symp Proc. 2012;2012:911-920.

29. Raja AS, Gupta A, Ip IK, Mills AM, Khorasani R. The use of decision support to measure documented adherence to a national imaging quality measure. Acad Radiol. 2014;21(3):378-383. doi:10.1016/j.acra.2013.10.017.

30. Lakshminarayan K, Rostambeigi N, Fuller CC, Peacock JM, Tsai AW. Impact of an electronic medical record-based clinical decision support tool for dysphagia screening on care quality. Stroke. 2012;43(12):3399-3401. doi:10.1161/STROKEAHA.112.662536.

31. Goldstein MK, Tu SW, Martins S, et al. Automating performance measures and clinical practice guidelines: differences and complementarities. AMIA Annu Symp Proc. 2014:37-38.

32. Richesson RL, Hammond WE, Nahm M, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. J Am Med Inform Assoc. 2013;20(e2):e226-31. doi:10.1136/amiajnl-2013-001926.

33. Persell SD, Kaiser D, Dolan NC, et al. Changes in performance after implementation of a multifaceted electronic-health-record-based quality improvement system. Med Care. 2011;49(2):117-125. doi:10.1097/MLR.0b013e318202913d.

34. Shelton JB, Skolarus TA, Ordin D, et al. Validating electronic cancer quality measures at Veterans Health Administration. Am J Manag Care. 2014;20(12):1041-1047.

35. Richesson RL, Smerek M. Electronic health records-based phenotyping. NIH Health Care Systems Research Collaboratory. http://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/. Published 2014. Accessed July 22, 2015.

36. Brown SH, Elkin PL, Rosenbloom ST, Fielstein E, Speroff T. eQuality for all: extending automated quality measurement of free text clinical narratives. AMIA Annu Symp Proc. January 2008:71-75.

37. Linder JA, Kaleba EO, Kmetik KS. Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. Med Care. 2009;47(2):208-216. doi:10.1097/MLR.0b013e318189375f.

38. Pakhomov S, Bjornsen S, Hanson P, Smith S. Quality performance measurement using the text of electronic medical records. Med Decis Making. 2008;28(4):462-470. doi:10.1177/0272989X08315253.

39. Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc. 2013;20(e1):e147-54. doi:10.1136/amiajnl-2012-000896.

40. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc. 2007;14(2):141-145. doi:10.1197/jamia.M2334.

41. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA. 1999;282(15):1458-1465.

42. Redwood S, Ngwenya NB, Hodson J, Ferner RE, Coleman JJ. Effects of a computerized feedback intervention on safety performance by junior doctors: results from a randomized mixed method study. BMC Med Inform Decis Mak. 2013;13(1):63. doi:10.1186/1472-6947-13-63.

43. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? Ann Intern Med. 2006;145(4):265-272.

CHAPTER 2

CHALLENGES AND RECOMMENDATIONS FOR INTEGRATION OF CLINICAL DECISION SUPPORT AND ELECTRONIC CLINICAL QUALITY MEASUREMENT: INSIGHTS FROM DOMAIN EXPERTS

2.1 Abstract

Objective of this study was to assess barriers and develop recommendations for the integration of clinical decision support (CDS) and electronic clinical quality measurement (eCQM).

Leading experts in CDS and eCQM were recruited using targeted invitations and an open solicitation on listservs for professional national informatics organizations. Through semistructured in-depth and critical incident interviews using online meeting software, we obtained stakeholders' insights about CDS and eCQM integration, with a focus on key differences and similarities between CDS and eCQM, benefits and barriers of integration, and potential solutions.

Fifteen experts were recruited, including executives and other leaders from academia, healthcare organizations, government, consulting companies, and commercial Health IT vendors. The experts identified multiple barriers to the integration of CDS and eCQM and offered advice for addressing those barriers, which the research team

synthesized into 10 recommendations. In particular, experts suggested improving the availability and adoption of standards, improving the approach to developing clinical practice guidelines and eCQM specifications, addressing cultural and structural differences between CDS and eCQM teams, and, finally, aligning financial reimbursement models with quality of care.

Integration of CDS and eCQM will likely require substantial effort including developing technical capabilities and changing organizational structures and cultures to align CDS and eCQM.

Integrating CDS and eCQM will require addressing several barriers. We anticipate that the expert insights elucidated in this study will facilitate CDS and eCQM integration and ultimately improvements in care quality and value.

2.2 Background

Clinical decision support (CDS) systems and electronic clinical quality measurement (eCQM) are 2 important computer-based strategies aimed at improving the quality of healthcare.[1] For the purposes of this study, we define CDS as the provision of pertinent knowledge and person-specific information to clinical decision makers to enhance health and healthcare.[2] Examples of CDS tools include alerts, order sets, care plans, protocols, documentation templates and tools, relevant data summaries, and dashboards. Such CDS tools can help clinicians provide evidence-based care for a specific individual or for a population of patients.[3]

In turn, we define eCQM as the measurement and tracking of the quality of healthcare services using electronic data. Clinical quality measurement results are used in

reports and feedback on clinician performance, accreditation reviews and institutional performance metrics. Clinical quality measurement (QM) is traditionally conducted using manual chart abstraction, but this domain is transitioning towards electronic data extraction. In conjunction with CDS, or on its own, eCQM can help to improve quality by providing feedback to relevant stakeholders.[4]

CDS and eCQM fundamentally address the same issue of identifying patients who should receive particular health or administrative interventions and determining whether they have received that intervention.[5–7] Coordination of vision, processes, and technologies, or *integration*, of CDS and eCQM domains has the potential to improve healthcare value.[8–10] CDS can facilitate the collection of data elements needed for the quality measures, and eCQM results can support iterative, data-driven refinement of the CDS. Other potential positive outcomes of integrating CDS and eCQM include reducing duplication of effort and minimizing inconsistencies in guidance recommendations.

Recognizing the importance of such integration, groups including the US federal government and the Institute of Medicine (IOM) are advising healthcare providers to tighten the feedback loop between CDS and eCQM. [11,12] An improved ability to establish such a virtuous feedback loop between quality improvement and continuous performance measurement is an important enabler for becoming a Learning Health Care System. Notably, the US Office of the National Coordinator (ONC) for Health Information Technology (IT) and the Centers for Medicare & Medicaid Services (CMS) have sponsored the public-private Clinical Quality Framework (CQF) initiative to harmonize health IT standards for CDS and eCQM to facilitate their integrated implementation.[13] Several standards and technological solutions have been suggested to enable integration

of CDS and eCQM.[14–19]

Despite this recognition of the importance of integrating CDS and eCQM, fully integrated quality improvement approaches are still rarely used and only sporadically reported in the literature. For example, only 3 out of 160 randomized clinical trials included in a review of CDS systems by Lobach et al. were accompanied by periodic performance feedback.[3] Additionally, when an integrated approach is used, it is often incomplete. For example, coordinated CDS ad eCQM efforts based on commercial EHR systems oftentimes use different tools for CDS and for eCQM implementations.[10,20] Moreover, family physicians report a lack of quality improvement infrastructure to co-deliver CDS and eCQM in their practices.[21] Finally, aspects of organizational culture and structure that inhibit integration of CDS and eCQM are poorly described in the literature. In summary, there is a need for research to better characterize how CDS and eCQM can be better integrated to improve care. To address this need, we sought insight from experts in the field to characterize the current state of CDS and eCQM integration and to identify potential approaches for advancing such integration moving forward.

2.2.1 Objective

This study aimed to explore the beliefs and perceptions regarding the integration of CDS and eCQM functionality and activities within healthcare organizations, using qualitative methods that engage subject matter experts (SMEs). Our specific objectives were to (1) describe similarities and differences in CDS and eCQM implementation and use, (2) describe potential benefits of the integration of CDS and eCQM, (3) describe technical and cultural barriers to integrating CDS and eCQM, and (4) formulate

recommendations for CDS-eCQM integration.

2.3 Materials and Methods

2.3.1 Study Design

A qualitative study was conducted using in-depth semistructured interviews with subject matter experts (SMEs). The critical incident technique was used during the interview process to identify components related to the key challenges in CDS and eCQM integration.

2.3.2 Research Team

The study was conducted by a multidisciplinary research team with experience in CDS, eCQM, clinical and public health informatics, standards-based interoperability, qualitative methods, cognitive task analysis, biostatistics, and information technology.

2.3.3 Subjects

SMEs were enrolled through an open invitation for participation made via email to relevant email listservs sponsored by the American Medical Informatics Association (AMIA) CDS and Implementation work groups, the Health Level 7 (HL7) Clinical Quality Information and CDS work groups, and the Clinical Quality Framework (CQF) initiative.[13] In order to maximize the representation of relevant expert insights, invitations were also sent to individuals identified as being key SMEs based on literature review and by the study personnel. Participation was open to all interested professionals in the field who had both of the following qualifications:

- Experience developing or using a quality measurement system, ***and***
- Experience developing or using CDS interventions.

Thirty individuals responded initially and 15 individuals decided to proceed with the interview. It has been previously shown that 12-13 interviews could be sufficient to gather a majority of insights.[22,23] Thus, we did not send any new invitations after conducting the 15 interviews.

At the beginning of each interview, a verbal consent was obtained for participating in the study, recording and transcribing the interview, and including participants' names in publications. A financial incentive ($40) was offered to the participants for their time, but some participants declined.

The study was approved by the University of Utah institutional review board (IRB) (Protocol # 00077948).

2.3.4 Interviews

One-hour in-depth semistructured interviews included 3 parts: (1) questions about the participants' background and experience with CDS and eCQM, (2) critical incident questions, and (3) general questions about integration of CDS and eCQM. We did not include a prespecified and constrained definition of the "integration" construct in the interview script in order to provide the respondent with flexibility to discuss any aspects of potential integration that they felt were important.

Questions about the participants' background and experience concerned their current organizational role, the type of organization, whether they had encountered CDS or eCQM first in their career, whether they had more experience with CDS or eCQM, and

the degree of integration between CDS and eCQM in their organization on a 1 to 10 scale.

The critical incident technique allows collecting rich data from the respondents' perspective and in their own words without forcing them into any given framework. The critical incident technique allows identifying even rare events that might be missed by other methods that only focus on common and everyday events.[24] The critical incident methodology was adapted from cognitive work analysis methods described by Crandall et al. where a 4-phase format was used: (1) incident identification, (2) timeline verification, (3) deepening, and (4) 'what-if' queries. First, we asked the interviewee to identify a specific project where he/she used both CDS and eCQM. Second, we asked the participant to provide a time-based description of the sequence of tasks in order to create an explicit timeline. Third, we asked a set of more specific questions to identify and verify project goals, social context, organizational issues, challenges, and decision points. Finally, a few "what-if" questions were posed to explore what could have been done differently under critical relevant conditions.

General questions about integration of CDS and eCQM included questions about similarities and differences between CDS and eCQM implementation and use, technical and nontechnical barriers to the integration of CDS and eCQM, and recommendations to reach a higher degree of integration.

Interviews were recorded and transcribed. All responses were used for the analysis.

2.3.5 Data Analysis

The interviews, including answers to critical incident questions and general questions about integration of CDS and eCQM, were analyzed using content analysis techniques described by Patton and Graneheim et al.[25,26] Transcript analysis began with one author (PK) identifying responses as relevant or not relevant using 5 predefined areas of interest as general categories.

The following taxonomy was chosen by study personnel for its pragmatic utility for understanding why CDS-eCQM integration is desirable, why such integration is still quite limited, and how integration could be achieved:

- similarities in CDS and eCQM implementation and use,
- differences in CDS and eCQM implementation and use,
- benefits to the integration of CDS and eCQM,
- technical and nontechnical barriers to the integration of CDS and eCQM, and
- recommendations for the integration of CDS and eCQM.

Relevant responses were reviewed at the paragraph level by a 3 person multidisciplinary research team with qualitative research experience. The team converted responses to condensed descriptions that preserved the meaning of the response. Then, related condensed descriptions and corresponding responses were summarized into constructs. The research team discussion was iterative, with condensed descriptions discussed, reviewed, and then reviewed again until no new constructs emerged within each area. Constructs were then aggregated into thematic statements within each area in order to elucidate the “gist” of the content. The thematic statements were then reviewed by the research team. Insights were not tied to any specific individual study participant.

Internal participant identification number is shown in parentheses after each quote.

2.4 Results

2.4.1 Participants

Fifteen SMEs with diverse backgrounds and organizational experience participated, including executives and other leaders from academia, government, healthcare provider organizations, consulting companies, and CDS and electronic health record (EHR) system vendors (Table 2.1).

Eleven SMEs first encountered CDS in their career. Among these 11 participants who encountered CDS first, 5 remained currently more experienced in CDS, 1 is now more experienced in QM, and 5 reported being equally experienced in both domains. In contrast, among the 4 participants who encountered QM first in their career, 3 remained currently more experienced with QM than CDS, while 1 of the 4 is now more experienced with CDS than QM.

When asked to report on current level of integration between CDS and eCQM on a scale of 1 to 10, SMEs varied widely in their responses. SMEs reported that optimal level of integration between CDS and eCQM has not been reached yet, even in most advanced healthcare systems. Three participants refrained from answering this question. One participant felt this question was only applicable to provider organizations.

Critical incident stories covered a wide range of use cases in different settings, including inpatient, outpatient, and emergency departments. Stories were told from different perspectives, including large healthcare systems and small practices, as well as consulting and vendor companies. The majority of responders described quality

improvement projects where both CDS and eCQM were used. For example, some projects were aimed at improving previsit planning reports for pediatric patients, or creating checklists to reduce cancellation rates at a cardiac surgery service. Goals of other projects included reducing hypoglycemic episodes, reducing catheter associated urinary tract infections, improving blood pressure control and diabetes management, improving timeliness of thromboembolism prophylaxis, prescribing warfarin for atrial fibrillation for patients that were high risk of stroke, and improving pneumonia management in emergency department. Some projects focused on implementation of eCQMs, such as for depression management, while other projects aimed to improve Medicare-related quality measures or the quality of clinical problem lists. Some of these projects succeeded and some failed according to the respondents' perceptions.

Over 250 quotations were extracted from the interview transcripts. The comments were summarized into condensed descriptions that describe similarities (n = 6 descriptions), differences (n = 21), benefits (n = 13), and barriers (n = 55). Additionally, comments related to potential solutions were summarized into 71 condensed descriptions, and then further summarized into 10 actionable recommendations. Sample responses, condensed descriptions, resulting constructs, and summarized thematic statements are presented in table format.

2.4.2 Similarities and Differences in CDS and eCQM Practice

All SMEs noted that CDS and eCQM are similar but also different in important aspects. Key similarities included the common goal of clinical quality improvement, the use of similar patient data for calculation, scalability requirements, and the need for logic

(Table 2.2a). Key differences in CDS and eCQM implementation and use, included differences in the level of analysis (ie, patient vs. population), whether eligible patients are defined strictly or loosely, and the culture and motivation of implementing teams (Table 2.2b).

2.4.3 Potential Benefits of Integration of CDS and eCQM

SMEs identified many potential benefits of integrating CDS and eCQM, including more effective quality improvement, better prioritization, and higher consistency of quality improvement interventions, reduced cost of implementation and financial benefits for the healthcare organization (Table 2.3). However, some SMEs were more optimistic than others about the potential to achieve those benefits. One participant also pointed that costs of integration of CDS and eCQM might outweigh benefits.

2.4.4 Barriers for Integrating CDS and eCQM

SMEs identified many technical and nontechnical barriers to implementing CDS and eCQM (Table 2.4). Five themes concerning barriers to integration were identified, including limited availability and adoption of standards and technological solutions, problems with authoring guidelines, different organizational cultural and structural barriers, and financial barriers.

2.4.5 Recommendations for Integration of CDS and eCQM

The SMEs noted that integration of CDS and eCQM will require contributions from many stakeholders, including standards developers, EHR vendors, CDS vendors,

eCQM vendors, CDS and eCQM implementers, healthcare executives, healthcare providers, guideline and quality measure authoring agencies, and the payer community. To accelerate integration of CDS and eCQM, 10 actionable recommendations were generated based on the insights of SMEs. The recommendations are grouped by stakeholder type.

- Standards Developers
    - Develop and improve harmonized standards, including standard terminologies, to represent executable logic, clinical data, and metadata that address both CDS and eCQM use cases.
- EHR Vendors
    - Develop EHR capabilities to coimplement CDS and eCQM. Provide ways to expose the data in a standard and secure way that can be used across both CDS and eCQM in a common manner.
- Technology Developers and Implementers
    - Use existing and emerging harmonized standards and technical approaches (eg, libraries of reusable elements and modules, data access standards) to implement and share CDS and eCQM knowledge across institutions.
    - Use a sustainable and robust maintenance strategy that includes versioning, documentation, validation, and updates to account for asynchronous changes in both CDS and eCQM specifications.
    - Use strategies for selecting evidence-based interventions and reconciling differences between CDS and eCQM definitions and requirements to account for multiple competing recommendations.

    - Engage all relevant stakeholders and iteratively develop common, streamlined solutions to account for the multidisciplinary nature of CDS-eCQM projects.
- Healthcare Executives and Organizational Leadership
    - Cross-train individuals who can serve as liaisons, develop coordinated governance, and create a culture of collaboration instead of competition to improve communication between CDS and eCQM groups.
- Guideline/Specification Authoring Groups
    - Specify corresponding CDS when developing eCQMs, and vice versa.
    - For eCQMs, use data elements already available in the EHR at the time of the encounter (eg, clinical data collected as a part of routine workflow) rather than depending on new documentation or billing data captured after the encounter.
- Payers and Government Agencies
    - Use financial incentives that promote CDS and eCQM, such as a "pay for value" reimbursement model.[27]

2.5 Discussion

Based on the insights from SMEs, CDS and eCQM integration could promote clinical improvement, increase consistency of quality improvement interventions, and reduce cost of implementation. However, they also described challenges that must be overcome before integration and subsequent efficiencies can be realized. These challenges include divergent standards and technical approaches, uncoordinated

specification authoring, lack of cultural and structural integration between CDS and eCQM teams in healthcare organizations, and misaligned financial incentives. CDS and eCQM have historically belonged to 2 different worlds within the healthcare enterprise. According to SMEs, CDS and eCQM professionals have different professional cultures and, oftentimes, have limited communication between each other. While many of these challenges are already currently being addressed, others remain outstanding and likely not fully appreciated by the stakeholders involved. To accelerate integration of CDS and eCQM, 10 actionable recommendations were synthesized based on the insights of SMEs in the fields of both CDS and quality measurement. In particular, the experts suggested improving availability and adoption of standards, changing the approach to CDS guidelines and eCQM specification development, addressing cultural and structural differences between CDS and eCQM teams, and aligning financial reimbursement models with quality of care.

Our study is different in scope and purpose from previously published manuscripts related to complementarities between CDS and eCQM. This study not only confirms and expands on the findings from previous studies with regard to similarities and differences between CDS and eCQM, our paper also describes challenges and provides recommendations for the integration.[5–7] Goldstein et al. described similarities and differences between CDS and eCQM with regard to cohort definitions, knowledge modeling, workflow integration, use of data, and output structures,[5] while Brown et al. compared CDS and eCQM in terms of data sources, analytic methods, units of analysis, delivery timing, intended users, and recommendations.[7] Haggstrom et al. focused on how the relationship between CDS and eCQM is perceived by relevant stakeholders. As in

these prior studies, this study found that CDS and eCQM are similar with regard to data sources but different in terms of analytical methods, units of analysis, delivery timing, cohort definitions, and intended users. The current study additionally found that CDS and eCQM differ in the professional cultures of the teams that implement these capabilities. The current study provides more detail compared to prior studies, with the inclusion of direct quotes from experts in the field to illustrate the many nuances of the complex relationship between CDS and eCQM. Furthermore, use of the critical incident technique allowed us to identify rare events such as conflicts and difficulties that may have not been reported otherwise.

The SMEs identified many potential benefits to integrate CDS and eCQM, including reducing costs, increasing alignment between CDS and eCQM implementations, and avoiding inefficient, duplicative efforts in each area. Other potential benefits identified include the coupling of CDS with automated performance feedback, improved quality, enhanced organizational efficiency, and financial benefits. Taken together, the integration of CDS and eCQM can help transform healthcare organizations into Learning Healthcare Systems with effective feedback loops for quality improvement.[11] However, as indicated by the wide variations in the provider responses about the degree of integration in their own organizations, there are large differences in the progress of organizations towards this goal. Furthermore, this variation could be partially explained by differences in participants' beliefs about what an ideal integration may entail.

Several efforts are underway to address the challenges to the integration of CDS and eCQM identified by the SMEs. For example, the CQF initiative sponsored by the

ONC and CMS is developing harmonized standards for data representation, metadata, and executable logic to facilitate coimplementation of CDS and eCQM.[13] Moreover, while many EHRs currently have limited native capabilities for coimplementation of CDS and eCQM, an evolving app marketplace may enable external vendors to produce standards-based solutions that could be used for both CDS and eCQM.[28] As for differences in professional culture, several promising projects are ongoing, including the development of knowledge centers in academic health systems that integrate CDS and quality measurement, such as the New York-Presbyterian Hospital's Value Institute and the Johns Hopkins Armstrong Institute for Patient Safety and Quality.[29,30]

While many challenges are already being addressed, others still need to be resolved. First, limited native EHR capabilities continue to be a problem, and EHR vendors are not necessarily prioritizing coimplementation of CDS and eCQM. Second, quality measure specifications generally do not include CDS guidance. More will need to be done with regard to the authoring of clinical guidelines and quality measures to facilitate the integration of CDS and eCQM. Third, cultural differences between teams and lack of coordinated governance, structure, and processes largely remain to be addressed. Indeed, SMEs mentioned that nontechnical barriers to CDS-eCQM integration are probably more important than the technical ones. Integration will need to be achieved at different levels, including for standards integration, IT infrastructure integration, specification authoring integration, and organizational and cultural integration.

If the CDS and eCQM stakeholders are able to address the described challenges, the vision of integrated and efficient quality improvement framework may be accomplished. We believe that the 10 recommendations provided in this paper can

facilitate this integration. Additionally, there may be 'game-changers' that facilitate this transition, including a focus on payment for value and the sponsorship of integration efforts by CMS, which can drive healthcare policy in the United States.

Our study has several potential limitations. First, as a qualitative study, the results may be influenced by the researchers' personal biases or by the phrasing of the interview questions. However, we used robust content analysis methodologies to help ensure the reliability of our findings.[25,26] We also include the interview script in the manuscript to make the questions available to the readers. Second, the inclusion and analysis of only 15 interviews may limit generalizability. Even though it has been previously shown that 12-13 interviews could be sufficient to gather a majority of insights,[22,23] more interviews may have provided more insights in this particular study. Third, the self-selection recruitment strategy may have biased the included SMEs to those who strongly agree or disagree that CDS and eCQM should be integrated. However, the resulting sample included SMEs representing a broad spectrum of healthcare professionals from many geographical regions and with different past experiences, enhancing our ability to describe the breadth of issues. Forth, our study does not include estimates of costs of CDS-eCQM integration. However, we felt that the qualitative nature of our research would not allow us to estimate whether benefits of integration outweigh cost. Thus, we decided to leave this topic out of scope. We therefore believe that our conclusions remain generalizable.

This study identified several areas where further research is needed to overcome remaining barriers to CDS and eCQM integration. In particular, there is a need to investigate strategies for mitigating the cultural differences and improving

communication between CDS and eCQM teams. In addition, there is a need to track progress and to evaluate the benefits and costs of enhanced CDS and eCQM integration through shared governance, infrastructure, and technical approaches.

2.6 Conclusion

This study improves our understanding of the challenges and opportunities for integrating CDS and eCQM. The findings could serve as a useful guide for ongoing activities in CDS-eCQM integration. Integration efforts will need to address many challenges, including those related to standards, technology, specification authoring, organization culture and structure, and financial incentives. While all the experts in the study agreed that integration of CDS and eCQM is important, the SMEs differed in their viewpoints on the feasibility of the integration in the near future. Integration of CDS and eCQM will likely require substantial effort for developing the necessary technical and organizational capabilities.

Table 2.1. Participants

Name of participant	Role/title	Name of Organization	Type of organization
Howard Bregman, MD, MS	Director, Clinical Informatics	Epic, WI	EHR vendor
Nathaniel Weiner, MS	Co-Founder, Chief Operating Officer	Avhana Health, MD	CDS vendor
Clement J. McDonald, MD	Scientific Director	US National Library of Medicine, MD	Government, academia*, healthcare provider*
Samson Tu, MS	Senior Research Scientist	Stanford University, CA	Government, healthcare provider
Art Wallace, MD, PhD	Chief, Anesthesia Service	San Francisco Veterans Affair Medical Center, CA	Government, healthcare provider
Adam Wright, PhD	Associate Professor of Medicine	Harvard Medical School, MA	Academia, healthcare provider
Keith Marsolo, PhD	Associate Professor	Cincinnati Children's Hospital Medical Center, OH	Academia, healthcare provider
Keith F. Woeltje, MD, PhD	Vice president, Chief medical informatics officer	BJC HealthCare, MO	Academia, healthcare provider
Hojjat Salmasian, MD, MPH, PhD	Program Director of Research Science	Value Institute, NewYork-Presbyterian Hospital, NY	Healthcare provider
Joseph Kunisch PhD, RN-BC, CPHQ	Enterprise Director for Clinical Quality Informatics- Regulatory Performance	Memorial Hermann Hospital System, TX	Healthcare provider
Benjamin Brown, MRCGP, MSc	General Practitioner and Research Training Fellow	University of Manchester, UK	Healthcare provider
Jerome A. Osheroff, MD	Founder/Principal	TMIT Consulting, LLC, TX	Consultant
Michelle Currie, RN, MSN, CPHQ, CPHIMS	Founder and Healthcare Solution Architect	Savant Solutions4HIT, LLC, CA	Consultant
William Salomon, MD, MS MPH	Senior Medical Informatician	Clinical Metrics, Limited liability company, ME	Consultant
James McCormack, PhD	Instructor - Health IT Project Management	Oregon Health & Science University, OR	Consultant

* - previous employment

Table 2.2. Similarities and differences between CDS and eCQM

Construct	Condensed Description	Sample Responses
2a. Similarities between CDS and eCQM		
Theme: CDS and eCQM aim to improve healthcare quality.		
Common goal	CDS and eCQM have the same purpose of healthcare quality improvement.	*"The clinical purpose is generally similar, in both cases. My goal, building an eCQM or building CDS, is to improve care …" (14)*
Theme: CDS and eCQM are based on similar patient data.		
Reliance on patient data	CDS and eCQM rely on similar patient data.	*"They're measuring the same thing, they're working on the same datasets, they are using electronic records, most of the time, or administrative data …" (15)*
Dependence on data quality	Results of CDS and eCQM are only as good as the quality of the underlying data.	*"They are both predicated on the quality of the electronic data, so they're only as good as the electronic data." (1)*
Theme: CDS and eCQM are automated approaches following similar logic and applied to large patient populations.		
Executable logic	CDS and eCQM are defined by a combination of logical expressions and value sets (eg, denominator criteria, numerator criteria); CDS and eCQM follow similar logic	*"There's a little clinical reminder that says, 'Hey, please do X, Y, Z,' and then you check to see how often people did X, Y, Z. And you can bug people to the point where they'll actually change their behavior." (5)*
Machine automation	CDS and eCQM require automation to be used at scale.	*"The similarity between them is that they're both obviously using technology to automate information." (11)*

Table 2.2 Continued

Construct	Condensed Description	Sample Responses
2b. Differences between CDS and eCQM		
Theme: eCQM is more retrospective and population based, with more conservative population definitions compared to CDS. Being retrospective, eCQM could use claims data. CDS is prospective, individual-centered, relies on current data, and has inclusive population definitions.		
Focus of analysis	CDS is generally more prospectively oriented, presented in real-time during the patient visit, and focused on changing clinician behavior and collecting data; eCQM tends to be more retrospective and is usually related to evaluation, monitoring, and developing a strategy to improve clinical quality.	*"There's also a difference in the temporality of it. Decision support typically occurs in real time or near real time, whereas quality measurement is usually after the fact, retrospective, looking back over a lot larger periods of time in the clinical data." (1)*
Data elements	CDS usually relies on the EHR data; eCQM could rely both on EHR data and on claims data, or even on manually abstracted data; eCQM can rely on 'future' data which are not available when CDS is firing, such as lab results, procedures completed, etc.	*"… So how do you then run decision support, when you're essentially required to consider data that hasn't even been recorded yet, right? … of course the coded diagnosis is not going to be generated for days to weeks after the clinical scenario that you're faced with. " (9)*
Level of evaluation	CDS is usually calculated at the patient level; eCQM could be aggregated at different levels; CDS is often triggered by a change in the patient data, such as a new problem; or by an action from the provider, such as opening the order entry dialog eCQM is usually run at periodic intervals, or on demand.	*"I mean eventually when you're doing CDS, you're basically taking the EMR and doing this at the patient level. Okay, if you're doing CQM you're doing this at the population level." (4)*
Population definitions	CDS may have more loose definitions since it is expected to cover all patients to whom the proposed definition might apply; eCQM may have more strict population definitions (with more exclusion criteria defined), to ensure appropriate comparisons over time or between organizations and benchmarks, especially if it is related to financial incentives.	*"Decision support is somewhat more crude in terms of how it's applied. … There's a lot of effort that goes into defining the population so that you're truly measuring what's important. I don't know if that same level of rigor yet exists on the decision support side." (3)*

Table 2.2 Continued

Construct	Condensed Description	Sample Responses
2b. Differences between CDS and eCQM		
Theme: CDS is context aware and should be integrated within clinical workflow, while eCQM is context independent.		
Context dependence	CDS is context aware; eCQM is context independent; CDS requires workflow integration, which could be associated with higher implementation effort; CDS might require clinician judgment.	*"I think in CDS, what's important is the context. … This person has this role. This is when the alert should appear. Clinical quality measurement only looks from a logical perspective." (10)*
Visualization and presentation	CDS and eCQM are usually presented differently given their different audiences and purposes; CDS is often presented in textual form, eg, as alerts, reminders, or smart forms; eCQM is often presented in a table, graph or dashboard.	*"Hypothetically, in an ideal world, you just have to define them once, and program all of the things once, and then just have two different visualizations for the data: one which happens at the point of care on a case-by-case basis and you want to send an alert out, and one which happens at population level on demand." (12)*
Theme: CDS and eCQM are implemented by different teams having different professional cultures and motivational factors.		
Professional culture	CDS tends to be implemented by IT and informatics teams; eCQM tends to be implemented by quality department specialists with analytics, public health, or nursing backgrounds.	*"I suppose just different cultures, the people who do the quality measurement tend to be more from a public health background or a nursing background whereas the people who do the CDS tend to be from an informatics or IT background, and so they don't always know exactly how they will work together." (14)*
Motivators	CDS efforts are often initiated from within the healthcare institutions and based on internal quality goals that can be locally defined; eCQM requirements are often externally regulated and incentivized, and evolve more slowly.	*"… a lot of our quality measures for better or worse right now come from the federal government or from an insurance company …" (14)*

EMR – Electronic Medical Record

Table 2.3. Benefits of an integrated approach to CDS and eCQM

Construct	Condensed Description	Sample Responses
Theme: Integration of CDS and eCQM will likely result in more effective quality improvement, better prioritization, and higher consistency of quality improvement interventions.		
Clinical improvement	More effective clinical improvement and adoption of evidence-based care; Facilitated implementation of the quality improvement cycle, including through baseline performance measurement and continuous tracking; Improved prioritization of quality improvement interventions; Improved documentation of contraindications, therapy, or discussion with the patient.	*"It takes, on average, about 17 years to get a doctor to implement a Level 1 standard of care. … Clinical decision support can be used to educate people about what to do and speed up this very prolonged timeframe. The clinical decision support speeds up the implementation of quality improvement and then you can use the system to see how well people are doing with it." (5)*
Improved consistency	Improved consistency of quality improvement interventions and recommendations.	*"It seems unfair that we would have inconsistencies between our CDS and our quality measures. I think like, we owe it to our users to harmonize those approaches." (14)*
Theme: Integration of CDS and eCQM will likely result in reduced cost of implementation and financial benefits for the healthcare organization.		
Improved efficiency by reducing time and cost of implementation	Reduced implementation burden and shorter production time within and across healthcare systems; Reusing approaches between CDS and eCQM; Improved data flow and data sharing within and between organizations, including commercial CDS and eCQM vendors; More robust system, where it is easier to fix errors.	*"It makes the production time incredibly shorter since you're working from a common set of concepts. You're basically creating your clinical content with an aim of doing CDS and quality measurement. Same set of concepts; therefore you're not having to basically worry about compatibility of different sets of content – meaning the CDS and the quality measurement being based on different things." (4)*
Financial benefits	Eligibility for government incentives and avoiding penalties; Opportunity to redirect eCQM funding to CDS development: there is currently significant funding from the federal government for eCQM related efforts, and this funding could be used to improve CDS as well.	*"The performance of an organization on eCQM, either an organization or an individual level, is probably tied to reimbursement some way. Or, if it's not tied today, it's going to be tied in the future. So organization would see a benefit to improving their scores in quality measurement. So therefore, they would want their CDS to be at least somewhat aligned with the quality measure performance." (9)*

Table 2.4. Barriers to the integration of CDS and eCQM

Construct	Condensed Description	Sample Quotes
Theme: Poor standards availability and adoption complicate development of advanced systems. Poor data quality complicates data transformation and utilization.		
Incomplete terminology and modeling standards	Incomplete standards, leading to inconsistent implementations; Not all clinical use cases supported by current standards.	*"… the standards don't support all of the use cases we would need…" (6)*
Poor standards adoption	Multiple unharmonized standards; Low standards adoption.	*"We've got automated processes to compute the measures, but all of those activities are not yet standards-based." (6)*
Poor data quality	Poor data quality inhibits integration between the systems.	*"We don't have all the data that we need in one system, it hasn't been validated …" (2)*
Theme: Currently existing technological solutions are suboptimal.		
Limited native EHR capabilities	Limited native EHR capabilities for coimplementation of CDS and eCQM, especially the cases with complex logic; Limited flexibility in EHR customization.	*"There's only so much you can do with clinical decision support without custom programming. … There are things we can think of but the EHR does not have the ability …" (3)*
Performance issues	Challenging optimization of algorithms, developed for individual patients, for thousands of patients at a time.	*"So, the performance issue of … how to efficiently convert it for applying these inclusion/exclusion criteria on the large cohort of the patients" (8)*
Diversity of platforms and unstable environment	Different EHRs and databases implemented in different health systems; Different software for CDS and eCQM; Fast pace of change in terminologies, standards, and EHR vendors.	*"Terminology changes; concepts change; standards for measurement change."* *"The issues of system integration in so far as performance measures and CDS are often built using different infrastructures" (4)*
Workflow issues	Invasive interventions; Interruptive data collection for eCQM; Not user friendly interfaces; Over-alerting clinicians; CDS not optimized for population management.	*"We are not that good yet at knowing when to show the CDS workflow or even less good about knowing when to show quality measures in the workflow. Right now, once a quarter, we send the quality report to your department chair and then they might meet with you and tell you what we're doing …" (14)*
Documentation, maintenance and versioning	Different expectations for provision of adequate documentation, maintenance and versioning; Higher expectations for CDS for timely roll-out and tracking of updates.	*"… you really have to maintain CDS and I think that's one of the harder problems with it. Medical care changes …" (5)*

Table 2.4 Continued

Construct	Condensed Description	Sample Quotes
Theme: eCQM and CDS content can diverge.		
eCQM guidelines do not always support CDS	eCQMs designed without thinking of CDS; eCQMs lacking a clear CDS counterpart; While CDS has to rely on currently available data, eCQM might need to use data that become available later.	*"When CQMs are developed by the committees, the expert panels that do them, and the stewarding organizations, they're not thinking in terms of CDS. …" (9)*
Conservative nature of many quality measures	eCQMs, particularly those used for compensation, may be more conservative than the care guidelines upon which CDS is based.	*"… But pay for performance measurement uses a target of 150 over 90, because the target of 140 over 90 is too difficult to reach: they don't want to interfere with payment." (15)*
Uncoordinated updating of CDS and eCQM specifications	Uncoordinated updating of quality measures and CDS guidelines, conducted on different timeframes; Prevalence of locally defined CDS interventions, as opposed to quality measures defined at a national level.	*"... measures are defined by Meaningful Use, by National Quality Forum … . And the CDS may be based on recommendations from professional societies … ." (8)*
Theme: Organizational and cultural factors inhibit integration of CDS and eCQM.		
Perception as separate domains	Perception of CDS and eCQM as two different approaches.	*"People just don't view these things as the same" (6)*
Cultural differences between teams	Cultural differences between CDS and eCQM teams; Difficulty communicating between IT, quality and other stakeholders with different worldviews.	*"So you have two ways of viewing the world, different terminology, and just different ways of talking and things like that." (6)*
Lack of coordinated governance, structure, and processes	Independent CDS and eCQM teams with limited processes for coordination; No organizational structure and governance for unified CDS and eCQM; Hard to get right people at the same table; EHR vendors have separate teams working on CDS and eCQM.	*"The CDS developers are creating tools that organizations use for CDS purposes. And they basically work to refine those tools and add new functionality. ... Whereas the CQM team is essentially trying to keep up with the regulatory requirements. And the end result is they don't have a lot of intersection …" (9)*
Competing interests	Independent groups with competing interests, each with desire to be the primary stakeholder in terms of decision making, resourcing, and recognition; Competing priorities.	*"I see a lot where a certain group wants to be the group that solves the problem, so that they can either get the recognition, or substantiate their position … ." (11)*

Table 2.4 Continued

Construct	Condensed Description	Sample Quotes
Not seeing a rational for integration	Preference for the tools people are most familiar with; Not seeing a benefit of integrating CDS and eCQM.	*"… tendency to believe in your tool. If you do CDS, that's because you think CDS is better, if you focus on quality measures, probably you think quality measurements are more population focused …" (14)*
Inadequate resources and training	Lack of informatics training of personnel; Limited IT resources.	*"… it may actually be that those standards exist and we just weren't aware of them, didn't know how to leverage them." (6)*
Theme: There is often no clear financial rationale to co-implement CDS and eCQM.		
No clear financial incentive for providers	Limited financial or clinical incentive for many providers to adopt eCQM-based CDS, coupled with potentially extra work	*"… the perception is there's no direct link between physicians clicking another thing and more money coming into the practice or better outcomes for the patient." (13)*
Need for funding	Limited funding for innovation	*"… That's the hardest: in any project it's getting the money." (5)*

2.7 References

1. Shojania KG, McDonald KM, Owens DK. *Evidence-Based Review Methodology for the Closing the Quality Gap Series (Vol. 1: Series Overview and Methodology)*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2004.

2. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. *J Am Med Inform Assoc*. 2007;14(2):141-145. doi:10.1197/jamia.M2334.

3. Lobach D, Sanders GD, Bright TJ, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. *Evid Rep Technol Assess (Full Rep)*. 2012;(203):1-784.

4. Anderson KM, Marsh CA, Flemming AC, Isenstein H RJ. *Quality Measurement Enabled by Health IT: Overview, Challenges, and Possibilities: An Environmental Snapshot*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.

5. Goldstein MK, Tu SW, Martins S, et al. Automating performance measures and clinical practice guidelines: differences and complementarities. *AMIA Annu Symp Proc*. 2014:37-38.

6. Haggstrom DA, Saleem JJ, Militello LG, Arbuckle N, Flanagan M, Doebbeling BN. Examining the relationship between clinical decision support and performance measurement. *AMIA Annu Symp Proc*. 2009;2009:223-227.

7. Brown B, Peek N, Buchan I. The case for conceptual and computable cross-fertilization between audit and feedback and clinical decision support. *Stud Health Technol Inform*. 2015;216:419-423.

8. Cleveringa FGW, Gorter KJ, van den Donk M, van Gijsel J, Rutten GEHM. Computerized decision support systems in primary care for type 2 diabetes patients only improve patients' outcomes when combined with feedback on performance and case management: a systematic review. *Diabetes Technol Ther*. 2013;15(2):180-192. doi:10.1089/dia.2012.0201.

9. Weber V, Bloom F, Pierdon S, Wood C. Employing the electronic health record to improve diabetes care: a multifaceted intervention in an integrated delivery system. *J Gen Intern Med*. 2008;23(4):379-382. doi:10.1007/s11606-007-0439-2.

10. Persell SD, Kaiser D, Dolan NC, et al. Changes in performance after implementation of a multifaceted electronic-health-record-based quality improvement system. *Med Care*. 2011;49(2):117-125. doi:10.1097/MLR.0b013e318202913d.

11. Institute of Medicine. *Leadership Commitments to Improve Value in Healthcare: Finding Common Ground*. Washington, DC: The National Academies Press; 2009.

12. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health

records. *N Engl J Med*. 2010;363(6):501-504. doi:10.1056/NEJMp1006114.

13. Kawamoto K, Hadley MJ, Oniki T, Skapik J. The clinical quality framework initiative to harmonize decision support and quality measurement standards: defined standards, pilot results, and moving beyond quality improvement. *AMIA Annu Symp Proc*. 2015:202-204.

14. Pathak J, Bailey KR, Beebe CE, et al. Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium. *J Am Med Inform Assoc*. 2013;20(e2):e341-8. doi:10.1136/amiajnl-2013-001939.

15. Kukhareva P V, Kawamoto K, Shields DE, et al. Clinical decision support-based quality measurement (CDS-QM) framework: prototype implementation, evaluation, and future directions. *AMIA Annu Symp Proc*. 2014;2014:825-834.

16. Li D, Endle CM, Murthy S, et al. Modeling and executing electronic health records driven phenotyping algorithms using the NQF Quality Data Model and JBoss® Drools Engine. *AMIA Annu Symp Proc*. 2012;2012:532-541.

17. Thompson WK, Rasmussen L, Pacheco JA, et al. An evaluation of the NQF quality data model for representing electronic health record driven phenotyping algorithms. *AMIA Annu Symp Proc*. 2012;2012:911-920.

18. Konrad R, Tulu B, Lawley M. Monitoring adherence to evidence-based practices: a method to utilize HL7 messages from hospital information systems. *Appl Clin Inform*. 2013;4(1):126-143. doi:10.4338/ACI-2012-06-RA-0026.

19. Raja AS, Gupta A, Ip IK, Mills AM, Khorasani R. The use of decision support to measure documented adherence to a national imaging quality measure. *Acad Radiol*. 2014;21(3):378-383. doi:10.1016/j.acra.2013.10.017.

20. Forrest CB, Fiks AG, Bailey LC, et al. Improving adherence to otitis media guidelines with clinical decision support and physician feedback. *Pediatrics*. 2013;131(4):e1071-81. doi:10.1542/peds.2012-1988.

21. Ivers N, Barnsley J, Upshur R, et al. "My approach to this job is...one person at a time": Perceived discordance between population-level quality targets and patient-centred care. *Can Fam physician Médecin Fam Can*. 2014;60(3):258-266. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3952764&tool=pmcentrez&rendertype=abstract. Accessed February 25, 2016.

22. Guest G. How many interviews are enough?: An experiment with data saturation and variability. *Field methods*. 2006;18(1):59-82. doi:10.1177/1525822X05279903.

23. Francis JJ, Johnston M, Robertson C, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychol Health*. 2010;25(10):1229-1245. doi:10.1080/08870440903194015.

24. Crandall B, Klein GA, Hoffman RR. *Working Minds. A Practitioner's Guide to Cognitive Task Analysis*. 1st ed. MIT Press: MIT Press; 2006.

25. Patton MQ. *Qualitative Research & Evaluation Methods*. 3rd ed. Thousand Oaks (CA): SAGE Publications; 2002.

26. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today*. 2004;24(2):105-112. doi:10.1016/j.nedt.2003.10.001.

27. Burwell SM. Setting value-based payment goals--HHS efforts to improve U.S. health care. *N Engl J Med*. 2015;372(10):897-899. doi:10.1056/NEJMp1500445.

28. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *J Am Med Informatics Assoc*. 2016;23(5):1-10. doi:10.1093/jamia/ocv189.

29. NewYork-Presbyterian Hospital's Value Institute. http://www.nyp.org/valueinstitute/. Accessed June 14, 2016.

30. Armstrong Institute for Patient Safety and Quality. http://www.hopkinsmedicine.org/armstrong_institute/. Accessed June 14, 2016.

CHAPTER 3

CLINICAL DECISION SUPPORT-BASED QUALITY MEASUREMENT (CDS-QM) FRAMEWORK: PROTOTYPE IMPLEMENTATION, EVALUATION, AND FUTURE DIRECTIONS

Reprinted from American Medical Informatics Association Annual Symposium proceedings, P. V Kukhareva, K. Kawamoto, D. E. Shields, D. T. Barfuss, A. M. Halley, T. J. Tippetts, P. B. Warner, B. E. Bray, and C. J. Staes, "Clinical Decision Support-based Quality Measurement (CDS-QM) Framework: Prototype Implementation, Evaluation, and Future Directions," pp. 825–34, 2014, Copyright (2014), with permission from American Medical Informatics Association.

Clinical Decision Support-based Quality Measurement (CDS-QM) Framework: Prototype Implementation, Evaluation, and Future Directions

Polina V Kukhareva, MPH[1], Kensaku Kawamoto, MD, PhD[1], David E Shields[1], Darryl T Barfuss[2], Anne M Halley[2], Tyler J Tippetts[1], Phillip B Warner, MS[1], Bruce E Bray, MD[1], Catherine J Staes, MPH, PhD[1]
[1]Department of Biomedical Informatics, University of Utah, and [2]Quality Improvement and Safety Department, University of Utah Healthcare, Salt Lake City, UT

Abstract

Electronic quality measurement (QM) and clinical decision support (CDS) are closely related but are typically implemented independently, resulting in significant duplication of effort. While it seems intuitive that technical approaches could be re-used across these two related use cases, such reuse is seldom reported in the literature, especially for standards-based approaches. Therefore, we evaluated the feasibility of using a standards-based CDS framework aligned with anticipated EHR certification criteria to implement electronic QM. The CDS-QM framework was used to automate a complex national quality measure (SCIP-VTE-2) at an academic healthcare system which had previously relied on time-consuming manual chart abstractions. Compared with 305 manually-reviewed reference cases, the recall of automated measurement was 100%. The precision was 96.3% (CI:92.6%-98.5%) for ascertaining the denominator and 96.2% (CI:92.3%-98.4%) for the numerator. We therefore validated that a standards-based CDS-QM framework can successfully enable automated QM, and we identified benefits and challenges with this approach.

Introduction

Overview of Clinical Quality Measurement (QM). Delivering quality healthcare is challenging due to ongoing and ubiquitous variation in health system processes that may lead to errors.[1] Measuring and reducing variation from evidence-based clinical best practices have been shown to improve quality and decrease costs of healthcare.[2] Despite multiple efforts undertaken to improve healthcare quality since the publication of the Institute of Medicine's reports entitled 'To Err Is Human'[3] and 'Crossing the Quality Chasm'[4], the quality of healthcare in the United States continues to be compromised by unnecessary variation in the implementation of clinical practice guidelines. Having a means to assess healthcare quality is essential for identifying deviations from evidence-based best practices and mitigating preventable errors.[3] Clinical quality measures are measures of processes, experience, and/or outcomes of patient care. There are increasing mandates and financial incentives to use electronic health records (EHRs) to measure quality as opposed to employing traditional manual processes for QM.[3,4] For example, the Meaningful Use (MU) recommendations issued by the federal Health Information Technology Policy Committee (HITPC) in November 2012 require the implementation of QM as well as decision support for high-priority conditions, and the use of related standards.[4,5] The practice and value of quality measurement has evolved over time. According to Meyer *et al.*, "in the last half century, the US has gone from defining quality, to measuring quality, to requiring providers to publicly report quality measures, and most recently, beginning to hold providers accountable for those results".[6] The National Quality Forum (NQF) was created as a public-private partnership to guide decisions regarding quality measure selection.[7] Until recently, quality measurement has relied mainly on the use of electronic claims data, manual chart abstraction, and patient surveys.[8] Currently, QM is required by public and private payers, regulators, accreditors and others that certify performance levels for consumers, patients and payers.[6] Current quality measurement systems in many hospitals include time-consuming manual paper and electronic record abstraction by a quality improvement specialist.[9,10]

At large academic medical centers such as University of Utah Health Care (UUHC), manual data abstraction is often followed by data analysis by an external organization such as the University HealthSystem Consortium.[11,12] University HealthSystem Consortium is an alliance of 120 academic medical centers and 299 of their affiliated hospitals representing academic medical centers with a focus on quality and safety excellence.[13,14] Manual chart abstraction at UUHC is performed by the Quality and Patient Safety Department, which has 28 employees, including 12 Quality Improvement Specialists.[15] There are several limitations with this process. For example, (a) three to six months may elapse between the time of a clinical procedure (e.g., a surgery) and the time when feedback

is given to a clinician; (b) human errors may be introduced during manual record review; and (c) only a subset of the clinical events is oftentimes selected for review, leading to gaps in quality assessment coverage.

Previous Work in Automating QM. One of the promises of implementing EHRs is the possibility for automatic generation of QM.[10] A MU-certified EHR must be able to export standardized quality reports, which can then "be fed into a calculation engine to compute various aggregate scores".[16] Following these recommendations, major EHR vendors such as Epic have started to integrate QM logic into their products.[17] Currently, however, only some EHR vendors offer quality measurements embedded in their system, and the scope of measures supported is not always comprehensive.[10,17] For example, KPHealth Connect has automated six of 13 Joint Commission measurement sets, and Epic has automated 44 NQF quality measures.[10,17] Vendor-based solutions may offer 'full sample' analysis, but the logic may be a 'black box.' Also, it is not always clear which version of each rule has been implemented or whether the quality measure logic is up-to-date. In addition, users may not have control over the logic to customize quality measurement. Even so, automated QM has the potential to provide quality reports on demand, may avoid human errors in manual abstraction, and can analyze 100% of patient encounters. Most ongoing efforts to produce automated quality measures are tied to a specific EHR system, and the executable logic for the quality measure is not sharable between different EHR systems.[10]

The Problem: Duplicative, Divergent Implementation of QM and CDS. It is intuitively obvious that CDS and QM are highly related, as QM focuses on who is eligible for a needed intervention (denominator identification) and who among them has received the needed intervention (numerator identification), whereas CDS focuses on who is eligible for a needed intervention and has not received the needed intervention (equivalent to numerator identification). However, to the best of our knowledge, there have been limited evaluation and validation in the literature of how technical approaches for one problem space can be re-used in the other, especially for standards-based approaches. This is important, because EHR certification criteria will likely drive much work in this field. It has been suggested that the two could be combined.[18] There have been efforts to combine CDS and QM logic, but the efforts were not standards-based and no conceptual framework was developed.[19,20]

Potential Solution: Leverage a Standards-based CDS Web Service across a Population for QM. Kawamoto *et al.* have previously suggested that a standards-based, service-oriented architecture could be used to make CDS logic sharable between different EHRs.[21] In this study, we hypothesized that this approach could be extended to encompass both CDS and QM given similarities in their functional requirements.

In pursuing this potential approach to CDS-based quality measurement (CDS-QM), a promising resource to leverage is an open-source, standards-based, service-oriented framework for CDS known as OpenCDS.[22] As shown in Figure 1, an EHR system can submit patient data to OpenCDS and obtain patient-specific assessments and recommendations that are provided to clinicians via alerts, reminders, or other CDS modalities.[23] OpenCDS is compliant with the HL7 Virtual Medical Record (vMR) and HL7 Decision Support Service (DSS) standards, and it leverages various open-source component resources, including the JBoss Drools knowledge management platform and Apelon Distributed Terminology System. Theoretically, then, OpenCDS could be used to measure quality as well as provide CDS. Moreover, the use of a CDS-based QM approach could potentially provide advantages for quality improvement compared to traditional approaches. Therefore, the objectives of this study were to: a) identify opportunities to enhance quality improvement using CDS-QM, b) design and implement a CDS-QM approach aligned with candidate CDS standards for Meaningful Use,[24] and c) evaluate the CDS-QM approach for a representative quality measure.

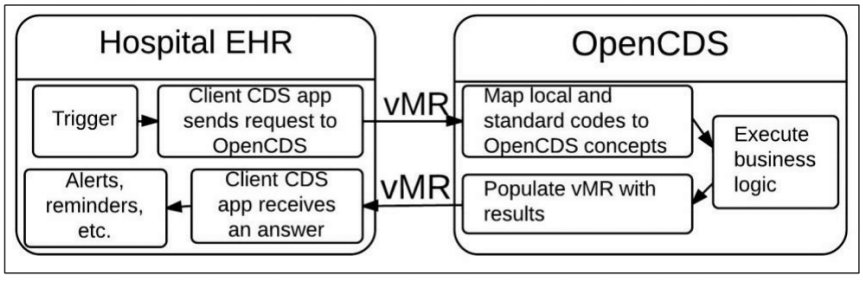


Figure 1. OpenCDS architecture: high-level interaction for CDS

Methods

Identification of opportunities to enhance quality improvement using CDS-QM

We engaged specialists from the Quality and Patient Safety Department to identify strengths and limitations of the CDS-QM approach. We documented the current process of quality assessment and reporting, and interviewed two of the 28 quality improvement specialists to identify possible ways in which the CDS-QM approach could enhance the institution's capabilities related to quality measurement and improvement.

Design and implementation of CDS-QM framework

Functional requirements

For the purposes of this implementation, our requirements were to enable the evaluation of national quality measures across the relevant patient populations in UUHC. The primary requirement was accurate evaluation of quality measure compliance.

Design principles

In designing the CDS-QM framework, a core design principle was *standards-based scalability*, so that the framework could potentially be leveraged in the context of various institutions and information systems. Related to this principle, a second principle was *open availability*, with open-source tooling used as to limit barriers to adoption related to licensing and intellectual property restrictions.

Scope and assumptions

Scope was limited to the analysis of structured data, as opposed to free text data requiring natural language processing. It was assumed that relevant patient data are available, such as in a data warehouse.

Tools and resources

In addition to OpenCDS, we leveraged the open-source Mirth Connect integration engine (v3.0.1). We also leveraged the UUHC data warehouse (DW), which contains data from the EHR systems and other ancillary clinical and administrative information systems at the institution.

Evaluation of CDS-QM approach for representative national quality measure

Quality measure

We chose the Joint Commission's Surgical Care Improvement Project (SCIP) Venous Thromboembolism 2 (SCIP-VTE-2) quality measure to evaluate the CDS-QM approach. This measure was chosen due to its technical complexity and its prioritization by the UUHC Quality and Patient Safety Department. VTE is a major cause of morbidity and mortality in hospitals.[25] In spite of evidence of their effectiveness, VTE prophylaxis by anti-coagulation and/or mechanical compression remains underutilized in US academic medical centers, particularly among surgical patients.[25] SCIP-VTE-2 is a well-established quality measure and is supported by level 1a evidence.[26,27] This measure is used to assess the percent of surgery patients that receive appropriate VTE prophylaxis within 24 hours of surgery.[28] For this evaluation, we implemented the SCIP-VTE-2 quality measure using the logic published for surgeries that occur in 2014 (version v4.2a).[28]

Data Sources

We used clinical data generated by inpatient surgeries that occurred at UUHC in 2013. A total of 8,924 cases were assessed using the CDS-QM automated method. The data elements required by the quality measure logic were documented in multiple source systems, including the inpatient EHR (Cerner) and two systems used to document anesthesia and nursing activities during the surgical event. The University of Utah Institutional Review Board performed an administrative review of this project and determined that IRB approval was not required because this effort was conducted for the purposes of quality improvement and does not meet the regulatory definition of human subject research.

Evaluation/Validation

As a reference standard for validation, we used quality measurement results produced by the University HealthSystem Consortium through an analysis of data extracted through manual chart abstraction by the Quality and Patient Safety Department. The sample used for validation was the 319 surgery cases randomly chosen for abstraction by the University HealthSystem Consortium for the first two quarters of 2013.

As the first step in our analysis, we evaluated the degree to which the data required for the evaluation was available in a structured format in the UUHC DW. Second, we compared the results from OpenCDS with the results from the reference standard approach (manual chart abstraction followed by University HealthSystem Consortium analysis). Observations were classified as true positive (TP), true negatives (TN), false positives (TP), and false negatives (FN) for denominator and numerator criteria separately. We calculated recall (sensitivity) of the OpenCDS-based process as the proportion of cases classified as positive by OpenCDS among the cases classified as positive by the reference standard (TP/(TP+FN)). We calculated precision (positive predictive value) of the OpenCDS-based process as the proportion of cases identified as positive by OpenCDS which was also classified as positive by the reference standard (TP/(TP+FP)). We assessed recall and precision for the classification of denominators, as well as recall and precision for the classification of numerators among cases that met denominator criteria. We also assessed the proportion of cases that yielded a complete match with the reference standard. Exact (Clopper-Pearson) confidence intervals (CI) were estimated for all binomial proportions. Statistical analysis was conducted using SAS version 9.3 (SAS Institute, Cary, North Carolina).

Results

Opportunities to enhance quality improvement using CDS-QM

As shown in Figure 2, the current quality assessment process at UUHC starts with data from the DW about clinical events (in this case major surgeries) being reported to the external quality benchmarking organization (the University HealthSystem Consortium), followed by this organization choosing a sample of surgery cases for manual chart abstraction. The UUHC quality specialists then perform manual abstraction of data from the EHR for the selected records. Finally, the UUHC specialist gives the information back to the external quality organization for summarization and reporting. The entire process from the time of surgery to quality reporting can take up to 6 months.

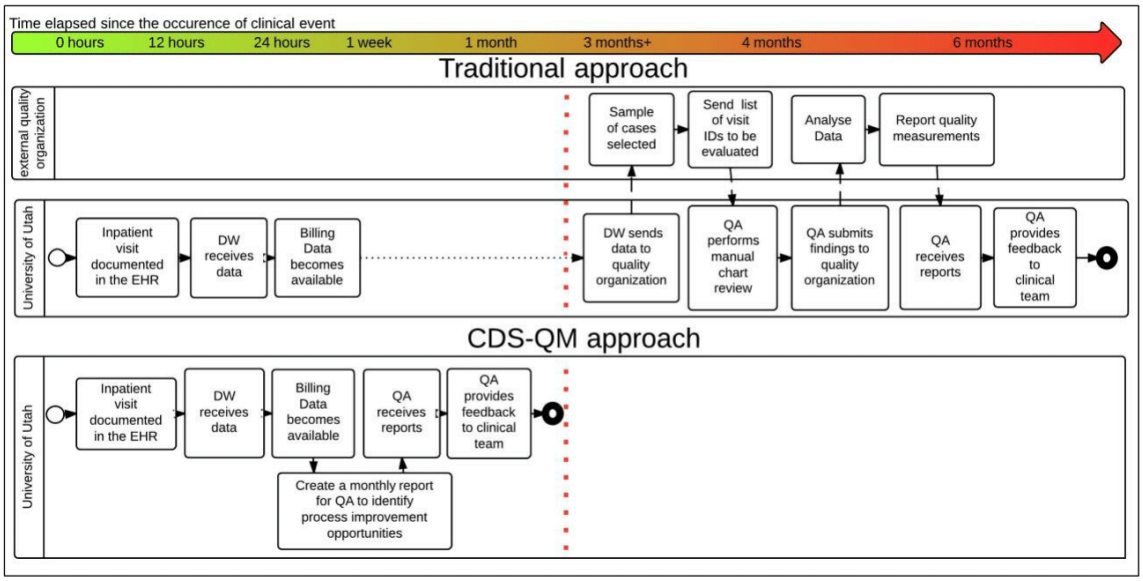


Figure 2. Comparison of traditional and CDS-QM approaches to quality measurement and improvement

Diagramming the traditional process for quality measurement and reporting was useful for identifying several opportunities using the CDS-QM approach to enhance workflow and impact clinical care. The informaticists and specialists with the Quality and Patient Safety Department identified several potential improvements. The automated approach using the CDS-QM strategy could:

- Improve the timeliness and completeness of feedback to the clinical stakeholders. The Quality Department holds monthly meetings with clinical stakeholders. Using the CDS-QM approach, a summary of the quality measure results from the previous month could be made available at these meetings to enable more rapid responses to identified quality deficiencies. The current reporting process allows such quality deficiencies to go unnoticed for potentially many months.
- Improve the completeness of assessment and feedback. While some records may still require manual review due to missing data (see Results), the vast majority of cases can be evaluated in an automated manner, as opposed to the baseline sampling approach. This more complete approach could potentially identify problem areas not yet sampled and therefore not yet identified by the quality team.
- Enable quality and clinical stakeholders to assess the impact of new rules or different versions of rules. This functionality is not available with the current process. At the same time, this functionality is critical for performing longitudinal analysis or for making predictions about future compliance.
- Provide additional useful information. A traditional quality measurement approach only identifies whether patients met denominator and/or numerator criteria. In contrast, the CDS-QM approach allows one to generate intermediate results that may be useful for better understanding the root cause underlying any deficiencies.

Design and implementation of CDS-QM framework

Figure 3 provides an overview of the CDS-QM approach. First, the Mirth Connect interface engine was used to identifying relevant surgery cases for analysis. Second, Mirth Connect was used to sequentially obtain relevant patient data from the DW. Third, the relevant patient data were converted into the HL7 vMR format used by OpenCDS. Then, Mirth Connect transmitted the vMR input data to OpenCDS using the SOAP Web service interface specified in the HL7 Decision Support Service (DSS) standard.

Within OpenCDS, the local and standard codes provided as the input were mapped to the internal concepts used by OpenCDS. The data were then evaluated by executing the relevant OpenCDS knowledge module, and the resulting patient-specific assessments were returned to Mirth Connect as vMR output objects. Finally, DW tables were populated with evaluation results by Mirth Connect.

Human workflows for developing the necessary business logic in Mirth Connect and OpenCDS were developed as well. These workflows involved identifying and characterizing the required input data, creating database queries in Mirth Connect, developing the required terminology mapping files, and developing the data processing algorithms in OpenCDS.

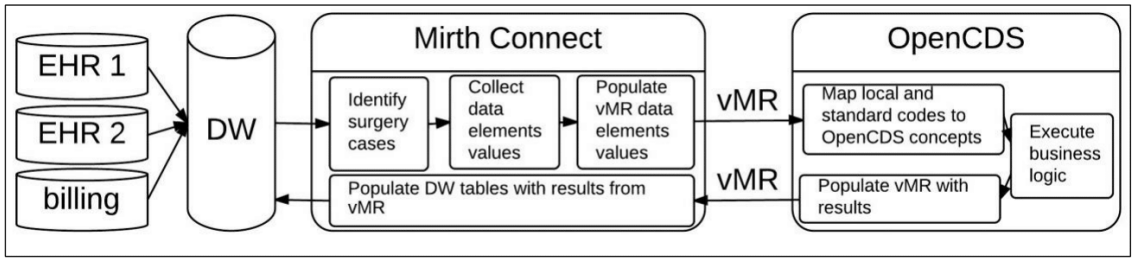


Figure 3. Major systems and processes involved in the CDS-QM approach

Evaluation of CDS-QM approach for representative national quality measure

The NQF-endorsed SCIP-VTE-2 quality measure was automated. Business logic was presented in the OpenCDS Web-based knowledge engineering platform, which uses the JBoss Guvnor platform. Figure 4 shows a sample screenshot from this knowledge engineering platform.

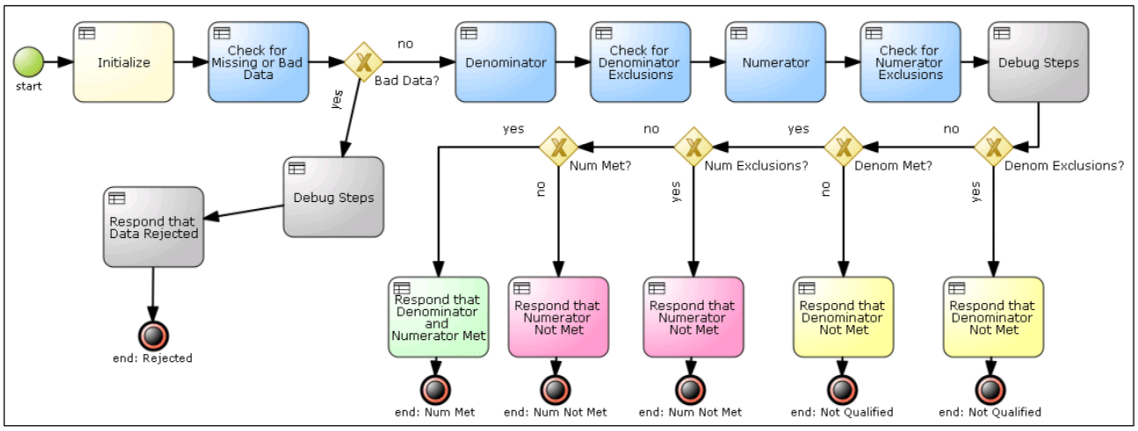


Figure 4. SCIP-VTE-2 Business Logic represented in Guvnor

Mirth queries were able to access all the necessary data elements except documentation about participation in clinical trials and the use of preadmission oral anticoagulant therapies. These two data elements were not recorded in the DW and would require manual review of the EHR text-based records for complete assessment. Among the 319 cases in the reference population selected by University HealthSystem Consortium, 14 did not have complete records stored in the DW and could not be used for automated QM. Completeness of the records was estimated as 95.6% (CI: 92.8%-97.6%). The remaining 305 cases were used to compare OpenCDS results with the reference standard.

When classifying denominators (i.e., identifying cases that should be included for quality measurement), the OpenCDS strategy yielded a recall of 100% (CI: 98%-100%) and precision of 96.3% (CI: 92.6%-98.5%) (see Table 1). A total of 183 cases among the 305 cases selected for review were found to meet the inclusion criteria. The seven cases included in the denominator using OpenCDS but excluded by the reference standard were cases with preadmission oral anticoagulant therapy. These cases were identified by the manual review process but not found by Mirth because the data is not currently saved in the DW.

Table 1. Comparison of the denominator (inclusion/exclusion) classification, using two methods (n=305)

CDS-QM approach (based on automated review using 2014 logic)	**Reference Standard** (based on manual review using 2013 logic)	
	include	exclude
include	TP=183	FP=7*
exclude	FN=0	TN=115

* had preadmission oral anticoagulants

Similarly, when classifying numerators (i.e., identifying cases that failed the quality measure among the 183 cases that met the denominator criteria), the OpenCDS strategy yielded a recall of 100% (CI: 97.9%-100%) and a precision of 96.2% (CI: 92.3%-98.4%) (see Table 2). The seven cases that passed following evaluation of the 2014 version of the SCIP-VTE-2 logic were cases that, in fact, failed according to the 2013 SCIP-VTE-2 specification version of the logic used for the manual review. These seven cases illustrate VTE-prophylaxis practices that were previously considered insufficient, but will be considered sufficient using the 2014 criteria. Thus, a complete match was found for 291 out of 319 records (91.2%; CI: 87.6%-94.1%).

Table 2. Comparison of the numerator (passed/failed) criteria, using two methods (n=183)

CDS-QM approach (based on automated review using 2014 logic)	**Reference Standard** (based on manual review using 2013 logic)	
	passed	failed
passed	TP=176	FP=7*
failed	FN=0	TN=0

* passed using 2014 logic

Discussion

We successfully prototyped an implementation of a CDS-QM-based system and demonstrated its feasibility. This use case demonstrates that, once the quality and CDS standards are fully aligned, the opportunity for meeting both goals using the same approach is quite feasible. We also identified special considerations for QM, such as the need to efficiently obtain and process data for a large cohort of patients, even while evaluation is conducted on a patient-by-patient basis.

As we engaged in the effort with the experts form the Quality and Patient Safety Department, we identified many opportunities to enhance quality improvement using CDS-QM. The currently used and proposed systems could complement one another. A CDS-QM system has the potential to be a cheaper and more efficient alternative to the analysis of quality measures relying on manual chart abstractions. However, it has a high initial cost for translating rules into executable format. Automation of the process of translating the measures into an executable format is a next logical step. Once translated into executable format, the logic can be shared with other institutions, and the logic can be modified to meet clinical needs to assess alternative or future quality measure specifications. The quality experts were particularly interested in the opportunity to apply the new SCIP VTE-2 logic specifications for 2014 to data generated from clinical practice occurring in 2013, which revealed that the new logic would reclassify their previous ‘failures’ as passing the new quality criteria. If the new logic had the opposite effect, it would be extremely helpful for a quality program to be able to anticipate ‘failures’ before they get reported six month later, as would occur using the traditional approach of manual abstraction and review by an external quality organization.

The CDS-QM approach is aligned with candidate CDS standards for Meaningful Use Stage 3. Standards proposed for 2015 voluntary EHR specification criteria were used[24]. We were able to use OpenCDS to implement quality measure logic published by the National Quality Forum, and we generated results that were either an exact match or could be explained where differences were observed. Many institutions use SQL queries instead of a CDS-QM approach because it requires less initial effort. However, maintaining an external rules repository is easier than when rules include a direct reference to the EHR or DW data, as there are no dependencies on the local database structure (i.e., the ‘curly braces problem’).

To maximize benefits from the use of a CDS-QM approach, high-quality structured data are necessary. Validation of the SCIP VTE-2 quality measure implementation using the CDS-QM approach highlighted gaps in documentation and problems with the transfer of information from source systems and the UUHC DW. These findings have been

shared with the DW and EHR teams, and the feedback is being used to improve processes and data being extracted into a surgery data mart. These feedback loops are important for improving data completeness and concordance. For data that was available in the DW, our automated approach shows favorable results compare to other automated approaches. For example, Kern *et al.* report that recall of electronic reporting ranged from 46% to 98% per measure, and precision from 57% to 97%.[29]

The CDS-QM approach may have limitations. Quality measures are often dependent on claims data, which are not usually available in real-time. More analysis is required to evaluate the timeliness of the availability of all the data required to implement quality measure logic. In addition, this study was performed in only one setting and focused on only one rule which may limit the generalizability of the results. However, OpenCDS uses a standard HL7 data model (the Virtual Medical Record), which potentially would allow the quality measurement rules to be implemented across systems. Additional research is necessary to demonstrate the CDC-QM approach in other EHRs and for other institutions.

In the future, we will automate more quality measures and provide additional feedback to the Quality and Patient Safety Department at UUHC, and the output metrics will be incorporated into dashboards that assess the cost and quality of care at UUHC. We are also working on modifying the OpenCDS infrastructure to support population-based queries. Instead of building one vMR at time, we are in the process of assembling thousands of vMRs at the same time. This approach to building vMRs should enable our approach to scale to quality measurement involving much larger patient samples, such as health maintenance measures for the general outpatient population.

Conclusion

In this study, we presented a prototype of the CDS-QM approach which can add value to the traditional quality reporting approaches. The CDS-QM approach allows for full case coverage, prospective use, near real-time evaluation, and is based on standards. The benefits of using CDS-QM include sampling a higher proportion of data, avoiding human error, saving abstractor time, providing more control over measurement rules, and independence from the EHR or other data source systems. To the best of our knowledge, this is the first study illustrating a framework and an approach for using an open-source, system-agnostic, standards-based CDS tool for continuous quality measurement.

Acknowledgements

This study was funded by University of Utah Health Care. The funding source played no role in the design and conduct of the study. KK is currently or recently served as a consultant on CDS to the Office of the National Coordinator for Health IT, ARUP Laboratories, McKesson InterQual, ESAC, Inc., Inflexxion, Inc., Intelligent Automation, Inc., Partners HealthCare, and the RAND Corporation. KK receives royalties for a Duke University-owned CDS technology for infectious disease management known as CustomID that he helped develop. KK was formerly a consultant for Religent, Inc. and a co-owner and consultant for Clinica Software, Inc., both of which provide commercial CDS services, including through use of a CDS technology known as SEBASTIAN that KK developed. KK no longer has a financial relationship with either Religent or Clinica Software. KK has no competing interests related to OpenCDS, which is freely available to the community as an open-source resource. All other authors declare no conflict of interest.

References

1. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med [Internet]*. 2010 Nov 25 [cited 2014 Feb 6];363(22):2124–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21105794
2. James BC, Savitz LA. How Intermountain trimmed health care costs through robust quality improvement efforts. *Health Aff (Millwood) [Internet]*. 2011 Jul [cited 2014 Jan 31];30(6):1185–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21596758
3. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Engl J Med [Internet]*. 2010 Aug 5 [cited 2014 Feb 6];363(6):501–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20647183
4. Blumenthal D. Launching HITECH. *N Engl J Med [Internet]*. 2010 Feb 4 [cited 2014 Feb 6];362(5):382–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20042745

5. Joseph AM. The stage 3 meaningful use preliminary recommendations: concerns are being raised. *MLO Med Lab Obs [Internet]*. 2013 Jul [cited 2014 Feb 20];45(7):64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24180080
6. Meyer GS, Nelson EC, Pryor DB, James B, Swensen SJ, Kaplan GS, et al. More quality measures versus measuring what matters: a call for balance and parsimony. *BMJ Qual Saf [Internet]*. 2012 Nov [cited 2014 Feb 13];21(11):964–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3594932&tool=pmcentrez&rendertype=abstract
7. Panzer RJ, Gitomer RS, Greene WH, Webster PR, Landry KR, Riccobono CA. Increasing demands for quality measurement. *JAMA [Internet]*. 2013 Nov 13 [cited 2014 Feb 12];310(18):1971–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24219953
8. Anderson KM, Marsh CA, Flemming AC, Isenstein H RJ. Quality Measurement Enabled By Health IT: Overview, Challenges, And Possibilities: An Environmental Snapshot [Internet]. 2012 [cited 2014 Feb 6]. Available from: http://healthit.ahrq.gov/sites/default/files/docs/page/final-hit-enabled-quality-measurement-snapshot.pdf
9. Silow-Carroll S, Edwards JN, Rodin D. Using electronic health records to improve quality and efficiency: the experiences of leading hospitals. *Issue Brief (Commonw Fund) [Internet]*. 2012 Jul [cited 2014 Feb 6];17:1–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22826903
10. Garrido T, Kumar S, Lekas J, Lindberg M, Kadiyala D, Whippy A, et al. e-Measures: insight into the challenges and opportunities of automating publicly reported quality measures. *J Am Med Inform Assoc [Internet]*. 2014 Jan 1 [cited 2014 Jan 24];21(1):181–4. Available from: http://jamia.bmj.com/content/21/1/181.long
11. University HealthSystem Consortium (UHC) [Internet]. [cited 2014 Jan 30];Available from: https://www.uhc.edu/
12. University of Utah Health Care - Salt Lake City, Utah [Internet]. [cited 2014 Jan 30];Available from: http://healthcare.utah.edu/about/
13. Xu R, Polk RE, Stencel L, Lowe DK, Guharoy R, Duggal RW, et al. Antibiogram compliance in University HealthSystem Consortium participating hospitals with Clinical and Laboratory Standards Institute guidelines. *Am J Health Syst Pharm [Internet]*. 2012 Apr 1 [cited 2014 Jan 29];69(7):598–606. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22441793
14. Together we stand, collaborating we excel: partnering to target common goals. A report on and abstracts of the UHC (University HealthSystem Consortium) 2009 Quality and Safety Fall Forum. October 1-2, 2009. Atlanta, Georgia, USA. *Am J Med Qual [Internet]*. [cited 2014 Jan 29];25(2 Suppl):20S–36S. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20220202
15. The University of Utah Basic Search - Campus Directory [Internet]. [cited 2014 Jan 30];Available from: http://people.utah.edu/uWho/basic.hml?did=2012
16. Dolin RH. Accuracy of electronically reported “meaningful use” clinical quality measures. *Ann Intern Med [Internet]*. 2013 Jul 2 [cited 2014 Jan 27];159(1):73. Available from: http://annals.org/article.aspx?articleid=1700652
17. Epic: Meaningful Use Stage 2 Certification Details [Internet]. [cited 2014 Feb 3];Available from: https://www.epic.com/software-certification.php
18. Lobach DF, Kawamoto K, Anstrom KJ, Kooy KR, Eisenstein EL, Silvey GM, et al. Proactive population health management in the context of a regional health information exchange using standards-based decision support. *AMIA Annu Symp Proc [Internet]*. 2007 Jan [cited 2014 Mar 3];:473–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2655792&tool=pmcentrez&rendertype=abstract
19. Raja AS, Gupta A, Ip IK, Mills AM, Khorasani R. The use of decision support to measure documented adherence to a national imaging quality measure. *Acad Radiol [Internet]*. 2014 Mar [cited 2014 Feb 13];21(3):378–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24507424
20. Lakshminarayan K, Rostambeigi N, Fuller CC, Peacock JM, Tsai AW. Impact of an electronic medical record-based clinical decision support tool for dysphagia screening on care quality. *Stroke [Internet]*. 2012 Dec [cited 2014 Jan 27];43(12):3399–401. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3508397&tool=pmcentrez&rendertype=abstract
21. Kawamoto K, Jacobs J, Welch BM, Huser V, Paterno MD, Del Fiol G, et al. Clinical information system services and capabilities desired for scalable, standards-based, service-oriented decision support: consensus assessment of the Health Level 7 clinical decision support Work Group. *AMIA Annu Symp Proc [Internet]*. 2012 Jan [cited 2014 Jan 22];2012:446–55. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3540445&tool=pmcentrez&rendertype=abstract

22. OpenCDS Home [Internet]. [cited 2014 Jan 30];Available from: http://www.opencds.org/
23. Kawamoto K, Shields D, Del Fiol G. OpenCDS: Enabling Clinical Decision Support at Scale through Open-Source, Standards-Based Software and Resources [Internet]. *AMIA poster.* 2011 [cited 2014 Feb 24];:1830. Available from: http://assets.conferencespot.org/fileserver/file/121836/filename/1lqa1c.pdf
24. Department of Health and Human Services (HHS). Voluntary 2015 Edition Electronic Health Record Certification Criteria: Interoperability Updates and Regulatory Improvements [Internet]. 2014 [cited 2014 Mar 7];Available from: http://www.regulations.gov/#!documentDetail;D=HHS-OS-2014-0002-0001
25. Schleyer AM, Schreuder AB, Jarman KM, Logerfo JP, Goss JR. Adherence to guideline-directed venous thromboembolism prophylaxis among medical and surgical inpatients at 33 academic medical centers in the United States. *Am J Med Qual [Internet]*. 2011 [cited 2014 Jan 29];26(3):174–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21490270
26. Turpie AGG, Lassen MR, Davidson BL, Bauer KA, Gent M, Kwong LM, et al. Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty (RECORD4): a randomised trial. *Lancet [Internet]*. 2009 May 16 [cited 2014 Jan 30];373(9676):1673–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19411100
27. Palmer AJ, Schramm W, Kirchhof B, Bergemann R. Low molecular weight heparin and unfractionated heparin for prevention of thrombo-embolism in general surgery: a meta-analysis of randomised clinical trials. *Haemostasis [Internet]*. [cited 2014 Jan 30];27(2):65–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9212354
28. National Quality Measures Clearinghouse | Surgical care improvement project: percent of surgery patients who received appropriate VTE prophylaxis within 24 hours prior to Anesthesia Start Time to 24 hours after Anesthesia End Time. [Internet]. [cited 2014 Jan 30];Available from: http://www.qualitymeasures.ahrq.gov/content.aspx?id=35538
29. Kern LM, Malhotra S, Barrón Y, Quaresimo J, Dhopeshwarkar R, Pichardo M, et al. Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study. *Ann Intern Med [Internet]*. 2013 Jan 15 [cited 2014 Feb 20];158(2):77–83. Available from: http://annals.org/article.aspx?articleid=1556360

CHAPTER 4

SINGLE-REVIEWER ELECTRONIC PHENOTYPING VALIDATION IN OPERATIONAL SETTINGS: COMPARISON OF STRATEGIES AND RECOMMENDATIONS

Reprinted from Journal of Biomedical Informatics, P. Kukhareva, C. Staes, K. W. Noonan, H. L. Mueller, P. Warner, D. E. Shields, H. Weeks, and K. Kawamoto, "Single-reviewer electronic phenotyping validation in operational settings: Comparison of strategies and recommendations," vol. 66, pp. Feb. 2017, Copyright (2017), with permission from Elsevier.

Journal of Biomedical Informatics 66 (2017) 1–10

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Single-reviewer electronic phenotyping validation in operational settings: Comparison of strategies and recommendations

Polina Kukhareva [a,*], Catherine Staes [a], Kevin W. Noonan [b], Heather L. Mueller [b], Phillip Warner [a], David E. Shields [a], Howard Weeks [b,c], Kensaku Kawamoto [a]

[a] Department of Biomedical Informatics and Knowledge Management and Mobilization, University of Utah, 421 Wakara Way, Suite #140, Salt Lake City, UT 84108, United States
[b] University of Utah Medical Group, 127 S. 500 E., Suite #660, Salt Lake City, UT 84102, United States
[c] Department of Psychiatry, University of Utah, 501 Chipeta Way, Salt Lake City, UT 84108, United States

ARTICLE INFO

Article history:
Received 23 June 2016
Revised 5 December 2016
Accepted 8 December 2016
Available online 9 December 2016

Keywords:
Electronic phenotyping
Computable phenotype
Validation
Electronic clinical quality measurement
Quality measure
Human chart review
Manual chart review

ABSTRACT

Objective: Develop evidence-based recommendations for single-reviewer validation of electronic phenotyping results in operational settings.
Material and methods: We conducted a randomized controlled study to evaluate whether electronic phenotyping results should be used to support manual chart review during single-reviewer electronic phenotyping validation (N = 3104). We evaluated the accuracy, duration and cost of manual chart review with and without the availability of electronic phenotyping results, including relevant patient-specific details. The cost of identification of an erroneous electronic phenotyping result was calculated based on the personnel time required for the initial chart review and subsequent adjudication of discrepancies between manual chart review results and electronic phenotype determinations.
Results: Providing electronic phenotyping results (vs not providing those results) was associated with improved overall accuracy of manual chart review (98.90% vs 92.46%, $p < 0.001$), decreased review duration per test case (62.43 vs 76.78 s, $p < 0.001$), and insignificantly reduced estimated marginal costs of identification of an erroneous electronic phenotyping result ($48.54 vs $63.56, $p = 0.16$). The agreement between chart review and electronic phenotyping results was higher when the phenotyping results were provided (Cohen's kappa 0.98 vs 0.88, $p < 0.001$). As a result, while accuracy improved when initial electronic phenotyping results were correct (99.74% vs 92.67%, N = 3049, $p < 0.001$), there was a trend towards decreased accuracy when initial electronic phenotyping results were erroneous (56.67% vs 80.00%, N = 55, $p = 0.07$). Electronic phenotyping results provided the greatest benefit for the accurate identification of rare exclusion criteria.
Discussion: Single-reviewer chart review of electronic phenotyping can be conducted more accurately, quickly, and at lower cost when supported by electronic phenotyping results. However, human reviewers tend to agree with electronic phenotyping results even when those results are wrong. Thus, the value of

Abbreviations: AAB, Avoidance of Antibiotic Treatment for Adults with Acute Bronchitis; AAP, Adults' Access to Preventive/Ambulatory Health Services; ABA, Adult Body Mass Index Assessment; AMM, Antidepressant Medication Management; AWC, Adolescent Well-Care Visits; BCS, Breast Cancer Screening; CAP, Children and Adolescents' Access to Primary Care Practitioners; CBP, Controlling High Blood Pressure; CCS, Cervical Cancer Screening; CDC, Comprehensive Diabetes Care; CDS, Clinical Decision Support; CHL, Chlamydia Screening in Women; CIS, Childhood Immunization Status; COL, Colorectal Cancer Screening; CWP, Appropriate Testing for Children with Pharyngitis; eCQM, electronic Clinical Quality Measurement; EHR, Electronic Health Record; eMERGE, Electronic Medical Records and Genomics; EMERSE, Electronic Medical Record Search Engine; FN, False Negative; FP, False Positive; FPC, Frequency of Ongoing Prenatal Care; HEDIS, Healthcare Effectiveness Data and Information Set; HPV, Human Papillomavirus Vaccine for Female Adolescents; IMA, Immunizations for Adolescents; LBP, Use of Imaging Studies for Low Back Pain; LSC, Lead Screening in Children; MPM, Annual Monitoring for Patients on Persistent Medications; NCS, Non-Recommended Cervical Cancer Screening in Adolescent Females; PBH, Persistence of Beta-Blocker Treatment After a Heart Attack; PPC, Prenatal and Postpartum Care; PPV, Positive predictive value; SHARPn, Strategic Health IT Advanced Research Project area four; TN, True Negative; TP, True Positive; URI, Appropriate Treatment for Children with Upper Respiratory Infection; UUHC, University of Utah Health Care; W15, Well-Child Visits in the First 15 Months of Life; W34, Well-Child Visits in the First 36 Years of Life.

* Corresponding author at: 421 Wakara Way, Ste 140, Salt Lake City, UT 84108-3514, United States.
E-mail addresses: polina.kukhareva@utah.edu (P. Kukhareva), catherine.staes@hsc.utah.edu (C. Staes), kevin.noonan@hsc.utah.edu (K.W. Noonan), heather.mueller@hsc.utah.edu (H.L. Mueller), phillip.warner@utah.edu (P. Warner), david.shields@utah.edu (D.E. Shields), howard.weeks@hsc.utah.edu (H. Weeks), kensaku.kawamoto@utah.edu (K. Kawamoto).

http://dx.doi.org/10.1016/j.jbi.2016.12.004

providing electronic phenotyping results depends on the accuracy of the underlying electronic phenotyping algorithm.
Conclusion: We recommend using a mix of phenotyping validation strategies, with the balance of strategies based on the anticipated electronic phenotyping error rate, the tolerance for missed electronic phenotyping errors, as well as the expertise, cost, and availability of personnel involved in chart review and discrepancy adjudication.

1. Introduction

Electronic phenotyping (i.e., automated identification of patients satisfying specified conditions) is essential for a variety of biomedical informatics domains including electronic clinical quality measurement (eCQM), clinical decision support (CDS), predictive analytics, risk adjustment, clinical registries, public health reporting, and cohort identification for clinical trials and research [1–3]. Moreover, the need for electronic phenotyping continues to increase due to the ongoing digitalization of health care [4]. For instance, many regulatory bodies are starting to require quality metrics to be calculated electronically instead of through manual chart audits. Given this increased need for electronic phenotyping, a number of projects have emerged for developing electronic phenotype definitions such as the Electronic Medical Records and Genomics (eMERGE), Electronic Medical Record Search Engine (EMERSE), mini-Sentinel, and Strategic Health IT Advanced Research area four (SHARPn) projects [5–7].

An electronic phenotype definition includes a set of inclusion and exclusion criteria that allow for the algorithmic selection of sets of individuals based on stored clinical data [1]. For example, a CDS system might recommend medications to improve glycemic control for individuals with diabetes who have hemoglobin A1c (HbA1c) levels of 9% or greater [8]. Similarly, an eCQM might require identifying the same set of individuals and determining whether those individuals received recommended care or achieved care goals. For a quality measure, denominator inclusion criteria refer to criteria specifying the set of all individuals for whom the measure is applicable (e.g., diagnosis of diabetes) [9]. Denominator exclusion criteria identify individuals who should be removed from the measure population before determining if numerator criteria are met (e.g., diagnosis of ischemic vascular disease). The numerator criteria are the processes or outcomes expected for each individual identified in the denominator (e.g., HbA1c < 7%). If denominator and numerator criteria are calculated using electronic algorithms, the determination of the patient status is referred to as an electronic phenotyping result [3].

Given the increasing importance of electronic phenotyping, it is essential that such phenotyping is as accurate as possible. For the purposes of this paper, we define an erroneous electronic phenotyping result as misclassification of a patient according to the phenotyping definition. For example, misclassifying an individual without diabetes as meeting denominator criteria for a comprehensive diabetes care measure is an inaccurate electronic phenotyping result. Such inaccurate phenotyping in the context of CDS can lead to alert fatigue and potentially patient harm [10], and inaccurate phenotyping in the context of eCQM can lead to misleading characterization of practice performance, and/or limit the ability of performance feedback to catalyze improvements in care quality [11,12]. Sources of error in electronic phenotyping include an incorrect or incomplete phenotyping definition, inaccurate interpretation of the phenotyping definition, erroneous identification or use of source clinical data, and insufficient data quality and consistency [13]. Many phenotyping algorithm implementations have been found to have problems with accuracy [14,15]. Therefore, electronic phenotyping results must be appropriately validated before they can be used with confidence. Electronic phenotyping validation is the process of establishing the accuracy of electronic phenotypes.

Electronic phenotyping validation requires comparing electronic phenotyping results to a reference standard. Such a reference standard is usually developed using one of the following three manual chart review methods: 'gold standard' (i.e., double manual chart review with at least two independent reviewers and adjudication, performed to resolve inter-reviewer discrepancies), 'trained standard' (i.e., one expert reviewer with validity of review checked), and 'regular practice' (single human reviewer) [16]. Among 113 studies of automated clinical coding and classification systems described by Stanfill et al., the majority (51%) used the single-reviewer 'regular practice' approach to create a reference standard [16]. While many phenotyping validation efforts use the 'gold standard' approach for iterative validation of phenotype definitions [1,5,17–19], such studies are usually performed in resource-rich research settings [6,13,19,20]. For example, the eMERGE network has conducted extensive research in phenotype definition validation such as identifying individuals with cataracts, type 2 diabetes, or dementia [13]. Unfortunately, 'gold standard' double manual chart review is often not feasible in operational settings due to resource constraints. Indeed, in operational settings, the reference standard for validation is often single manual chart review coupled with expert adjudication of any discrepancies with electronic phenotyping results.

Despite the importance of the single human reviewer approach in operational settings, this approach to electronic phenotyping validation is not well described in the literature. Thus, we have a limited understanding of the strengths and limitations of different single-reviewer strategies. In particular, there is limited guidance available in the literature on how a maximum number of phenotyping errors can be identified with minimal person-hours, thereby optimizing the quality of electronic phenotyping results given available resources.

At our academic medical center, we were faced with the need to efficiently validate electronic phenotypes being implemented for enterprise clinical quality measurement and physician compensation using a single-reviewer approach. It has been previously shown that providing electronic phenotyping results to humans can improve the efficiency of manual phenotyping tasks such as diagnosis coding and quality measurement [21,22]. However, no prior literature was available in the context of single-reviewer electronic phenotyping validation. Thus, we hypothesized that supporting a single-reviewer chart review process with electronic phenotyping results would make the review faster and more precise by reducing the validator's cognitive load. For example, providing the date of the last retinal eye exam for an individual with diabetes would allow the reviewer to more efficiently confirm that the patient received the recommended care. However, we were concerned that the reviewer may be influenced by the provided results and may over-agree with erroneous electronic phenotyping

results. Therefore, we conducted a randomized controlled trial in which the human reviewer performed chart audits both with and without the availability of the electronic phenotyping results. In so doing, we sought to develop evidence-based recommendations for single-reviewer validation in operational settings, with a particular focus on whether electronic phenotyping results should be provided to human reviewers during the validation process.

2. Objectives

The goal of our study was to develop evidence-based recommendations for single-reviewer validation of electronic phenotyping results in operational settings. Our study objectives were to (1) evaluate the impact of providing electronic phenotyping results on the quality and efficiency of manual chart review and validation and (2) describe human errors that occur when performing manual chart review with and without the support of electronic phenotyping results.

3. Materials and methods

3.1. Design

A randomized, controlled study was performed to evaluate the effects of providing results of electronic phenotyping on single-reviewer manual chart review effectiveness.

3.2. Setting

This study was performed at University of Utah Health Care (UUHC), an academic healthcare enterprise comprised of four hospitals, 10 community clinics, and several specialty centers [23]. UUHC currently uses the Epic electronic health record (EHR) system enterprise-wide.

3.3. Study population

The study population included patients who were enrolled in selected health plans or had at least one encounter at UUHC within one year prior to January 1, 2015 (N = 407,823 patients). The electronic phenotyping algorithms were applied to the full clinical population. Manual chart audits were conducted for 3528 test cases randomly selected from the study population stratifying by the results of the electronic phenotyping. Ten test cases were excluded from further analysis because historical vaccination data altering their phenotypes were entered in the medical record following the validation process. To eliminate imbalance, we randomly restricted the sample for each quality measure to have the same number of test cases per stratum across the two review strategies (Table 1). Statistical analysis was performed using a sample of 3104 test cases (88% of all chart-reviewed cases).

3.4. Intervention

We provided the reviewer with a spreadsheet containing a list of test cases to be evaluated (Fig. 1). In the control review strategy, reviews were conducted without the provision of electronic phenotyping results (Fig. 1A). In the intervention review strategy, reviews were conducted with the provision of electronic phenotyping results, including the intermediate details (e.g., date of the last colonoscopy, dates of immunizations, and date of the last blood pressure measurement) and final conclusions (numerator and denominator statuses) (Fig. 1B).

3.5. Manual chart review

Single-reviewer manual chart reviews to assess the accuracy of the electronic phenotyping results were performed by an analyst with 16 years of experience conducting phenotyping validation for billing, coding and quality measurement (HM). The reviewer manually analyzed the EHR patient data and classified the patient's status with regard to the denominator and numerator criteria. Chart reviews were completed as a part of the reviewer's routine work responsibilities.

3.6. Electronic phenotyping

At UUHC, quality measurement has been implemented using a CDS-based approach to eCQM [24] that leverages an open-source, standards-based phenotyping platform known as OpenCDS [25]. For each quality measure, final results (denominator and numerator statuses) are evaluated for each patient (Fig. 2). Then, based on a combination of inclusion and exclusion criteria, patients are classified into one of four strata (Fig. 2). This phenotyping framework also computes intermediate findings such as the dates of relevant encounter diagnoses and the counts of relevant medication prescriptions. These intermediate findings can be used to support debugging, facilitate manual chart review, and investigate discrepancies between chart reviews and electronic phenotyping results.

3.7. Quality measures

Twenty-six groups of quality measures defined by the National Committee for Quality Assurance's Healthcare Effectiveness Data and Information Set (HEDIS), version 2015, were implemented and validated [26]. These measure groups are presented in Table 1. Each HEDIS measure group includes one to 11 measures. For example, the CDC measure group includes a measure for the proportion of patients with diabetes with an HbA1c level less than 8% and a measure for the proportion of patients with diabetes with an eye exam performed in an appropriate timeframe. A total of 59 measures were implemented.

This research was partially motivated by our earlier experience with implementing the 2014 version of the HEDIS specifications, with the exception of those in the HPV and CBP measure groups. The validation process for these earlier measures used an ad hoc methodology: validation case selection was left up to the reviewer rather than being systematic, balanced, and random; we did not blind the reviewer to our available electronic phenotyping results; and we only had available the final electronic findings (i.e., numerator and denominator statuses).

3.8. Validation framework

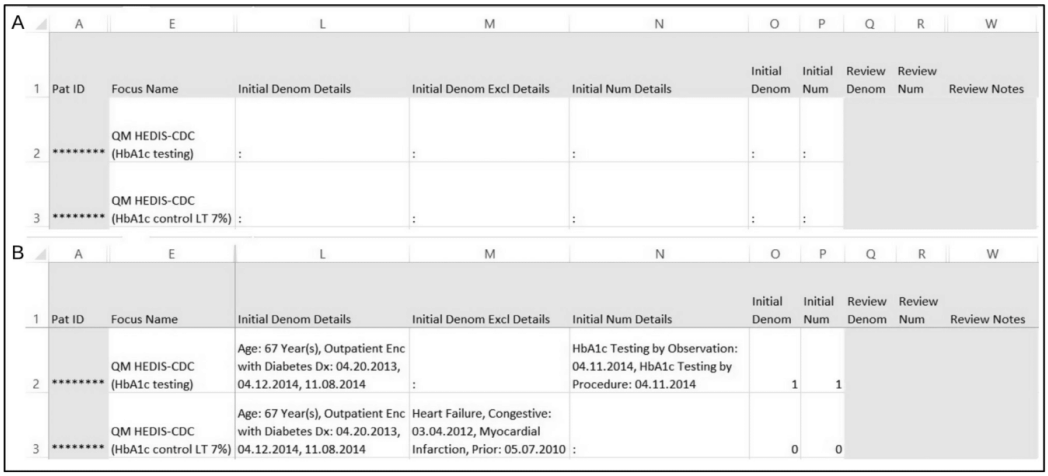
In order to replace the ad hoc validation procedures and tools we had used earlier, we developed a new standard approach to validation. The validation framework was implemented using the open source Groovy v2.4 programming language [27]. Clinical data were accessed from the UUHC data warehouse, an Oracle database that contains clinical and administrative data from the EHR and ancillary systems.

The validation framework enabled (1) the application to choose a representative random sample of test cases stratified by the electronic phenotyping results; (2) the reviewer to document her findings and comments as well as the time required for review; and (3) the discrepancy adjudication team to document its findings and comments. The application consists of the following software components:

Table 1
Quality measure groups and chart review sample size for each stratum.

		Measure group	N^*_m	No. of test cases used per review strategy							
				Control (N = 1552) Strata				Intervention (N = 1552) Strata			
				1	2	3	4	1	2	3	4
1	AAB	Avoidance of Antibiotic Treatment for Adults with Acute Bronchitis	1	10	0	10	10	10	0	10	10
2	AAP	Adults' Access to Preventive/Ambulatory Health Services	1	10	0	10	10	10	0	10	10
3	ABA	Adult Body Mass Index Assessment	1	9	0	10	10	9	0	10	10
4	AMM	Antidepressant Medication Management	2	20	0	17	18	20	0	17	18
5	AWC	Adolescent Well-Care Visits	1	10	0	10	10	10	0	10	10
6	BCS	Breast Cancer Screening	1	5	2	10	10	5	2	10	10
7	CAP	Children and Adolescents' Access to Primary Care Practitioners	1	10	0	10	10	10	0	10	10
8	CBP	Controlling High Blood Pressure	1	10	0	10	10	10	0	10	10
9	CCS	Cervical Cancer Screening	1	2	7	10	10	2	7	10	10
10	CDC	Comprehensive Diabetes Care	7	50	5	56	58	50	5	56	58
11	CHL	Chlamydia Screening in Women	1	10	0	10	10	10	0	10	10
12	CIS	Childhood Immunization Status	11	88	0	89	80	88	0	89	80
13	COL	Colorectal Cancer Screening	1	3	4	10	10	3	4	10	10
14	CWP	Appropriate Testing for Children with Pharyngitis	1	10	0	10	10	10	0	10	10
15	FPC	Frequency of Ongoing Prenatal Care	5	45	0	43	43	45	0	43	43
16	HPV	Human Papillomavirus Vaccine for Female Adolescents	1	10	0	10	10	10	0	10	10
17	IMA	Immunizations for Adolescents	3	30	0	24	28	30	0	24	28
18	LBP	Use of Imaging Studies for Low Back Pain	1	4	6	10	10	4	6	10	10
19	LSC	Lead Screening in Children	1	10	0	10	10	10	0	10	10
20	MPM	Annual Monitoring for Patients on Persistent Medications	4	36	0	38	33	36	0	38	33
21	NCS	Non-Recommended Cervical Cancer Screening in Adolescent Females	1	5	4	10	10	5	4	10	10
22	PBH	Persistence of Beta-Blocker Treatment After a Heart Attack	1	4	5	10	6	4	5	10	6
23	PPC	Prenatal and Postpartum Care	2	18	0	17	17	18	0	17	17
24	URI	Appropriate Treatment for Children with Upper Respiratory Infection	1	4	5	10	10	4	5	10	10
25	W15	Well-Child Visits in the First 15 Months of Life	7	56	0	57	61	56	0	57	61
26	W34	Well-Child Visits in the First 3–6 Years of Life	1	10	0	10	10	10	0	10	10
Total			59	479	38	521	514	479	38	521	514

N^*_m - No. of individual quality measures.



A

	A	E	L	M	N	O	P	Q	R	W
1	Pat ID	Focus Name	Initial Denom Details	Initial Denom Excl Details	Initial Num Details	Initial Denom	Initial Num	Review Denom	Review Num	Review Notes
2	********	QM HEDIS-CDC (HbA1c testing)	:	:	:	:	:			
3	********	QM HEDIS-CDC (HbA1c control LT 7%)	:	:	:	:	:			

B

	A	E	L	M	N	O	P	Q	R	W
1	Pat ID	Focus Name	Initial Denom Details	Initial Denom Excl Details	Initial Num Details	Initial Denom	Initial Num	Review Denom	Review Num	Review Notes
2	********	QM HEDIS-CDC (HbA1c testing)	Age: 67 Year(s), Outpatient Enc with Diabetes Dx: 04.20.2013, 04.12.2014, 11.08.2014	:	HbA1c Testing by Observation: 04.11.2014, HbA1c Testing by Procedure: 04.11.2014	1	1			
3	********	QM HEDIS-CDC (HbA1c control LT 7%)	Age: 67 Year(s), Outpatient Enc with Diabetes Dx: 04.20.2013, 04.12.2014, 11.08.2014	Heart Failure, Congestive: 03.04.2012, Myocardial Infarction, Prior: 05.07.2010	:	0	0			

Fig. 1. A. Screenshot of a spreadsheet for control review strategy. B. Screenshot of a spreadsheet for intervention review strategy. Dates have been randomly offset to anonymize patient data.

- The *Sampler* draws a stratified random sample of test cases into a spreadsheet (Fig. 1). Stratification was according to the four electronic phenotyping strata.
- The *Matcher* compares manual chart review results with electronic phenotyping results.
- The *Uploader* records findings in the application database.

We estimated the resources required for implementing and maintaining this validation framework based on retrospective review. The cost involved a clinical informatician (PK) building the application for sampling, matching and uploading results. The application was built on top of the existing electronic phenotyping infrastructure. The cost of adding intermediate results to

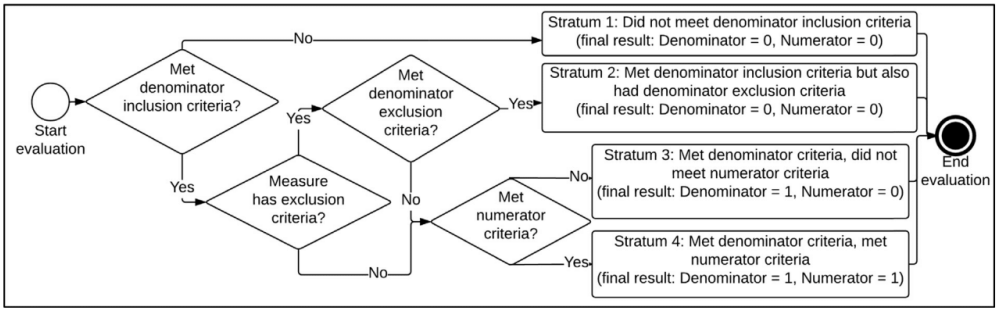


Fig. 2. Quality measure evaluation process.

the phenotyping framework was not included in this cost estimation for the validation framework because this feature was added to facilitate routine debugging.

3.9. Validation process and development of the reference standard

The reference standard was defined as the final phenotype determinations following the correction of all issues identified through the validation process.

Using the Sampler, we randomized 3528 test cases to control and intervention review strategies (Fig. 3). For quality measure groups containing more than one measure, test cases from the same patients were used across measures to facilitate chart review.

Using the Matcher, we compared manual chart review results with electronic phenotyping results for each test case (Fig. 3). If the manual chart review results matched the electronic phenotyping results, the reference standard was set automatically, without additional review. If there was a discrepancy between electronic phenotyping results and manual chart review results in any round of the regression testing, both manual chart review results and electronic phenotyping results were reviewed by the adjudication team.

Discrepancy adjudication was conducted by two content developers, with a faculty-level physician informaticist (KK) participating in the adjudication of difficult cases. The adjudicators used information from the EHR and the data warehouse, with additional input from the intermediate results, to come to consensus about the true patient phenotype. The adjudication team was not blinded to the review strategy. Conclusions were discussed and validated with the original reviewer (HM). Errors in electronic phenotyping were analyzed to find and fix root causes.

Whenever any changes were made to the phenotyping system, regression testing was conducted to compare the new phenotyping results against the previously validated reference standard. When an error in the electronic phenotyping process was discovered, fixing that error typically fixed all errors due to the same cause. The final reference standard was used to evaluate the accuracy of both the manual chart review results and the electronic phenotyping results.

3.10. Statistical analysis

3.10.1. Outcomes

To evaluate the performance of the chart review strategies the following dependent variables were calculated:

- Accuracy: percentage of test cases correctly classified by the human reviewer in comparison to the reference standard (N = 3104). Accuracy was estimated overall and for each stratum separately.
- Accuracy given correct initial electronic results: percentage of test cases correctly classified by the human reviewer among the test cases where the initial electronic phenotyping results were correct in comparison to the reference standard (N = 3049).
- Accuracy given erroneous initial electronic results: percentage of test cases correctly classified by the human reviewer among the test cases where the initial electronic phenotyping results were erroneous according to the reference standard (N = 55).
- Average time to review a test case (seconds): the review duration per test case per measure. The review duration was available for all measure groups except for LSC, where the information was inadvertently not recorded by the reviewer.
- Cohen's kappa: a measure of the inter-rater agreement between the results from manual chart review and electronic phenotyping. We combined strata 1 and 2 for this analysis.

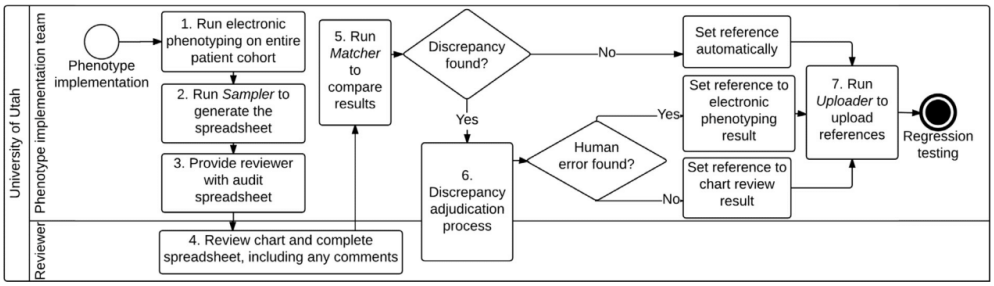


Fig. 3. Validation process and development of the reference standard.

Electronic phenotyping accuracy (i.e., the proportion of correct electronic phenotyping results) was estimated within each review strategy to make sure that it did not differ between the two groups.

Unadjusted tests (t-tests and chi-square tests) were used to compare the dependent variables between the review strategies. Generalized linear models were fitted to adjust for the patient stratum and correlation of responses from the same patient using generalized estimating equations in the genmod procedure (SAS statistical software).

The marginal cost of detecting one erroneous electronic phenotyping result was evaluated as follows:

- Number of erroneous electronic phenotyping results to review was estimated as 1/(accuracy of manual chart review when electronic phenotyping result is wrong).
- Overall number of test cases to review was estimated as (number of erroneous electronic phenotyping results to review)/(-electronic phenotyping error rate).
- Number of discrepancies to investigate was estimated as ((number of test cases to review − number of erroneous electronic phenotyping results to review) ∗ (probability of human error given correct electronic phenotyping results) + 1).
- Person-hours were estimated as (number of test cases to review ∗ time to review a test case in seconds)/(60 ∗ 60).
- Cost was estimated as number of reviewers ∗ person-hours ∗ pay rate per hour.

Based on our validation environment and estimates of typical personnel costs, we assumed the following:

- observed electronic phenotyping error rate;
- two adjudicators per discrepancy resolution;
- adjudication duration equal to average initial review duration in the control group;
- average chart reviewer pay rate of $25 per hour; and
- adjudication specialist pay rate of $100 per hour.

To evaluate if the difference in cost estimates between the review strategies was statistically significant, we adapted the "rearranging test" simulation methodology from Chang [28]. Specifically, we rearranged test cases between the study groups 10,000 times under the null-hypothesis, kept the same number of test cases in each stratum in each measure, and calculated what percentage of differences between estimates was below the actual difference value from the sample.

P-values less than 0.05 were considered significant. Statistical analyses were performed using SAS statistical software, version 9.3 (Cary, North Carolina).

3.10.2. Description of human errors with manual chart review

We categorized the reasons for the human errors to understand the limitations of manual chart review. The analysis was stratified according to whether the initial electronic phenotyping result was correct or erroneous.

The University of Utah Institutional Review Board approved this study (protocol 00081641).

4. Results

4.1. Accuracy of manual chart review

We validated 26 quality measure groups using 3104 test cases. Providing electronic phenotyping results (vs not providing the results) was associated with improved overall accuracy of manual chart review (98.90% vs 92.46%, p < 0.001) (Table 2). There was a difference in review accuracy within the strata. The control strategy showed poor performance in stratum 2 (where both denominator inclusion and exclusion criteria were met), with an average accuracy of 60.53%. While accuracy improved when given correct electronic phenotyping results (99.74% vs 92.67%, N = 3049, p < 0.001), there was a trend towards decreased accuracy when given erroneous electronic phenotyping results (56.67% vs 80.00%, N = 55, p = 0.07).

Across the 26 HEDIS quality measure groups, improvements in manual chart review accuracy varied (Fig. 4A). Providing electronic phenotyping results was associated with an over 15% improvement in accuracy for three quality measure groups: MPM, CAP and CCS.

After adjusting for the patient status and correlation among results from the same patient, parametric analysis showed that providing electronic phenotyping results was associated with an 83% reduction (95% confidence interval (CI): 70–90%, p < 0.001) in the odds of a human error, and a 95% reduction (95% CI: 87–98%, p < 0.001) in the odds of a human error when given correct electronic phenotyping results.

4.2. Duration of manual chart review

Providing electronic phenotyping results decreased the review duration per test case (62.43 vs 76.78 s, p < 0.001) (Table 2). The availability of electronic phenotyping results reduced the time to review test cases by more than 40% for four quality measure groups: AWC, NCS, URI and W34 (Fig. 4B). These four measures require a reviewer to find encounters for specific visit types (e.g., well care visit with a primary care provider), and this information was provided in the intermediate electronic phenotyping results.

Table 2
Performance of intervention and control strategies.

Performance	Review strategy		p-value[b]
	Control (N = 1552)	Intervention (N = 1552)	
Accuracy (N = 3104)	92.46%	98.90%	<0.001
Accuracy, Stratum 1[a] (N = 958)	98.54%	100.0%	0.008
Accuracy, Stratum 2[a] (N = 76)	60.53%	100.0%	<0.001
Accuracy, Stratum 3[a] (N = 1042)	90.21%	98.27%	<0.001
Accuracy, Stratum 4[a] (N = 1028)	91.44%	98.44%	<0.001
Accuracy given correct electronic results (N = 3049)	92.67%	99.74%	<0.001
Accuracy given erroneous electronic results (N = 55)	80.00%	56.67%	0.07
Accuracy of electronic phenotyping algorithm	98.39%	98.07%	0.5
Average time to review a test case (seconds)	76.78 +/- 90.02	62.43 +/- 80.70	<0.001
Cohen's kappa	0.88	0.98	<0.001
Estimated resource use to detect one erroneous electronic phenotyping result			
N of erroneous electronic results to review	1.25	1.76	0.08
N of test cases to review	70.55	99.59	0.08
Chart review person-hours	1.50	1.73	0.48
Cost of chart review	$37.62	$43.18	0.48
N of discrepancies to investigate	6.08	1.26	<0.001
Discrepancies adjudication person-hours	0.26	0.05	<0.001
Cost of discrepancies adjudication	$25.95	$5.36	<0.001
Cost to detect one erroneous electronic result	$63.56	$48.54	0.16

[a] Stratum 1: In initial electronic evaluation, test case did not meet denominator inclusion criteria.Stratum 2: In initial electronic evaluation, test case met denominator inclusion criteria but also had denominator exclusion criteria.Stratum 3: In initial electronic evaluation, test case met denominator criteria but did not meet numerator criteria.Stratum 4: In initial electronic evaluation, test case met denominator criteria and met numerator criteria.

[b] P-values computed using chi-square and rearrangement tests.

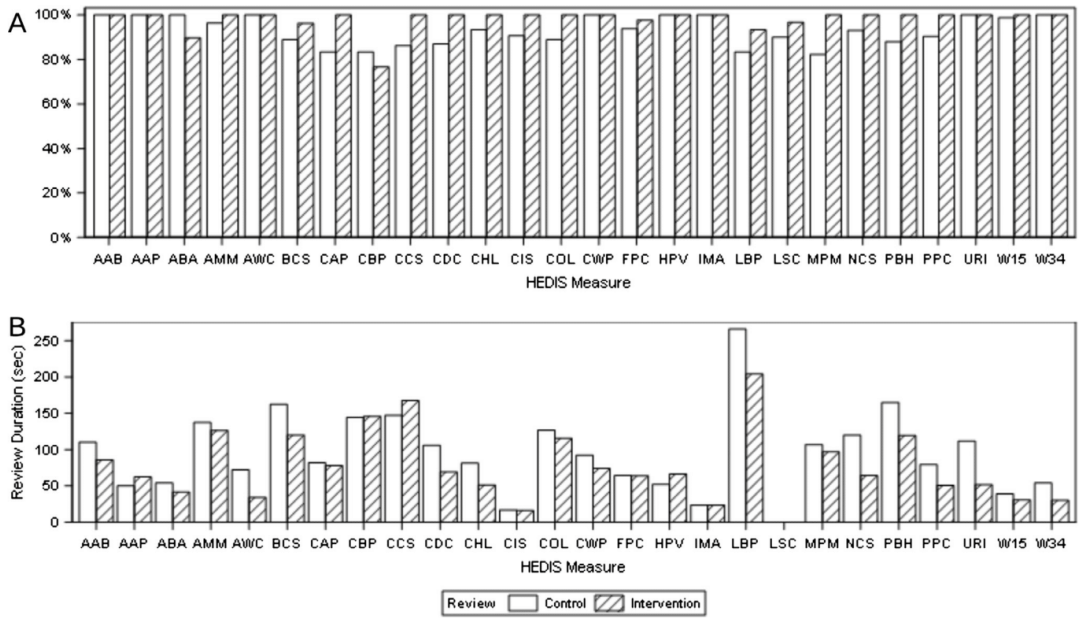


Fig. 4. Impact of the availability of electronic phenotyping results across 26 HEDIS quality measure groups on: (A) average accuracy of review and (B) average time to review a test case.

After adjusting for the patient status and correlation among results from the same patient, parametric analysis showed that providing electronic phenotyping results was associated with a 20% faster review (95% CI: 10–29%, $p < 0.001$).

4.3. Cost of detecting one erroneous electronic phenotyping result

The intervention review strategy insignificantly reduced the estimated marginal costs of identifying an erroneous electronic phenotyping result compared to the control strategy ($48.54 vs $63.56, $p = 0.16$). Based on the observed accuracy of the manual chart review and the average time to review a test case, we estimate that to find one electronic phenotyping error, 99.59 test cases would need to be reviewed if electronic phenotyping results are available during the chart review, and 70.55 test cases would need to be reviewed if electronic phenotyping results are not available (Table 2). However, since the review was faster in the intervention group, the overall time to identify one true electronic phenotyping error was similar (1.5 vs 1.73 h, $p = 0.48$). Additionally, the intervention group review generated fewer erroneous results when the electronic phenotyping results were correct, which saved resources required to investigate discrepancies. As a result, the marginal cost of identifying an erroneous electronic phenotyping result was approximately $48.54 for the intervention group and $63.56 for the control group.

We estimated the cost of implementing the electronic phenotyping validation infrastructure to be $2000. After the applications were implemented, the maintenance cost within the study period was minimal.

4.4. Ability to detect errors in electronic phenotyping results

Using both the initial validation process and subsequent regression testing, we found 55 electronic phenotyping errors resulting from 13 root causes (Table 3). These 55 erroneous results were discovered iteratively. Initially, the control strategy resulted in the discovery of 20 out of 25 (80%) of the electronic phenotyping errors, while the intervention strategy resulted in the discovery of only 17 out of 30 (56.67%) electronic phenotyping errors ($N = 55$, $p = 0.07$). In other words, while the control strategy missed 20% of erroneous electronic results, the intervention strategy missed 43.33%.

4.5. Description of human errors with manual chart review

Numerous reasons were identified for human errors in the chart review process (Table 4). Table 4 summarizes human errors when electronic phenotyping results were correct, and when electronic phenotyping results were erroneous.

The majority of human errors were caused by failure to find (i.e., missing) the relevant clinical statements (e.g., encounters, procedures, medication dispensation events, and diagnoses). Other human errors were caused by miscounting encounters, medication dispensation events, and medication days' supply in cases where the count of multiple events was relevant.

5. Discussion

Electronic phenotyping requires efficient and effective validation. However, there is no widely accepted standard process for validating single-reviewer electronic phenotyping results in operational settings. To our knowledge, this is the first study to provide evidence-based recommendations based on a systematic analysis of human errors associated with validating results of electronic phenotyping using a single-reviewer manual chart review process.

We developed a toolset and standard procedures to enhance the electronic phenotyping validation process in an operational setting. Compared to the ad hoc validation procedures and tools we used earlier, the new approach provided significant benefits, including (a) systematic, balanced, and randomized case selection;

Table 3
Ability of manual chart review to detect errors in electronic phenotyping for control and intervention chart review strategies.

Reason for electronic phenotyping error (N = 55)	Affected measure groups	Error detection rate[a] Review strategy	
		Control (N = 25)	Intervention (N = 30)
Electronic phenotyping errors in logic			
A wrong value set was used to identify outpatient visits	AAP	0/0	1/1
Patients without hypertension diagnoses were included due to an error in logic implementation	CBP	4/5	1/8
Patients' blood pressure was compared to an incorrect range due to an error in logic implementation	CBP	1/1	0/0
Patients with only one diabetes emergency department visit were included in the denominator due to an error in the logic implementation	CDC	1/1	9/9
Duplicate vaccine administrations were counted due to an error in the logic implementation	HPV	0/0	2/2
Electronic phenotyping errors with data quality			
Data warehouse did not include all body mass index values available in the EHR	ABA	1/1	1/3
Information about the end date of a medication order was updated in the EHR after the initial run.	AMM	0/0	1/1
Reversed version of the combination drug name was missed by the matching algorithm	CDC	1/1	0/0
Values of gestational age were unavailable for some patients due to an incorrect data query	FPC	6/6	0/0
Some provider specialties were incorrectly mapped	FPC, PPC	2/6	0/3
Some tetanus, diphtheria, and pertussis (Tdap) vaccination events were not mapped correctly	IMA	3/3	2/2
External lead screening tests were unavailable for some patients due to an incorrect data query	LSC	0/0	0/1
Some emergency department visits were counted as outpatient visits due to an incorrect data query	W15	1/1	0/0

[a] Electronic phenotyping errors detected during review/all errors in the original electronic phenotyping.

(b) workflow and data management support, including for the initial review, discrepancy adjudication, and regression testing; and (c) the ability to rigorously evaluate the impact of making intermediate and final electronic phenotyping results available to human reviewers. Finally, the approach could potentially be adopted by other institutions to enable the efficient and systematic validation of electronic phenotyping results in operational settings.

The quality and efficiency with performing manual chart review is influenced by the availability of intermediate and final results generated by electronic phenotyping. The intervention and control review strategies (i.e., providing or not providing electronic phenotyping results) each have strengths and limitations. The strengths of providing electronic phenotyping results to a reviewer include faster review, higher overall accuracy, higher accuracy when electronic phenotyping results are correct, and fewer resources required for discrepancy adjudication. This finding is supported by Garrido et al., who showed that providing automated results during the manual chart review process reduced abstraction time by 50% for Joint Commission quality measures [21]. However, Garrido et al. assumed that their electronic phenotyping process was 100% correct, which may not always be true.

According to our findings, a limitation of the intervention review strategy is that humans may agree with incorrect electronic phenotyping results in many cases. Therefore, the intervention review strategy may result in electronic phenotyping errors being overlooked. Missing electronic phenotyping errors may have many negative consequences such as inaccurate recommendations and alert fatigue in the CDS domain and inaccurate quality measurement results in the eCQM domain. At the same time, however, the intervention review strategy enables conducting more chart reviews for the same amount of resources, thereby increasing the total number of electronic phenotyping errors that can be identified given the same amount of resources.

In contrast, strengths of the control review strategy are that it eliminates the influence of potentially incorrect results and can increase the identification of true electronic phenotyping errors. Unfortunately, a limitation is that this strategy can produce a large number of human errors that can result in unnecessary, resource-intensive adjudication.

While manual chart review is usually considered a reference standard when assessing the quality of electronic phenotyping results, our findings show that manual chart review is subject to error and may be more erroneous than electronic phenotyping. In our study, the accuracy of manual chart review (92.47%) was lower than the accuracy of electronic phenotyping (98.39%) when using the control review strategy. This finding supports well-established notions of the 'imperfectability of man' as a processor of complex logic [29]. Indeed, many of the HEDIS measures used in this study have complex logic processing requirements. For example, the HEDIS FPC measure group specifies over a dozen algorithms for identifying appropriate prenatal care visits, which require the simultaneous consideration of such factors as the provider specialty, the time since the estimated conception, associated diagnoses, and associated procedures. In addition, finding rare exclusion criteria can be challenging for humans, especially when a patient has many comorbidities and a large medical record. Indeed, many human errors come from stratum 2 where reviewers identified common inclusion criteria but overlooked less common exclusion criteria (Table 2). Therefore, just as studies have shown that clinician performance can be improved through clinical decision support [30], this study adds to a body of evidence showing that the performance of human reviewers can be improved through the provision of relevant electronic phenotyping data, but only if the electronic data are correct.

This study has several strengths and limitations. With regard to strengths, we used a randomized controlled study design to evaluate the impact of providing intermediate and final electronic phenotype data to a human reviewer. To our knowledge, this is the first study to use this type of a rigorous study design to evaluate the pros and cons of making such information available in the context of single-reviewer electronic phenotyping validation. As a second strength, we had a robust sample of over 3000 chart reviews conducted across more than 20 phenotype groups. Finally, we conducted an economic analysis to identify the potential financial implications of the two strategies for conducting single-reviewer phenotyping validations in operational settings.

An important limitation is that we did not conduct double manual chart review to establish the reference standard. Thus, if the human reviewer agreed with the electronic phenotyping result, the results was simply assumed correct, even though both results may have been wrong. However, we sought in this study to specifically evaluate how single-reviewer validation processes could be optimized in operational settings. As another limitation, this study was conducted in a single institution and for a single type of electronic phenotyping (HEDIS quality measures). Therefore, further

Table 4
Types and examples of human errors in manual chart review results.

Reason for human error	Review strategy		Examples
	Control	Intervention	
Human errors when electronic phenotyping results were correct (N = 116)			
Missed clinical statements for inclusion criteria	43	0	Reviewer missed major depression diagnosis, encounters with diabetes medications and diagnoses, pregnancy test, and prescriptions for angiotensin converting enzyme inhibitors, diuretics and digoxin. Reviewer also failed to include patient with the right age
Missed clinical statements for numerator criteria	16	2	Reviewer missed: administered vaccinations, colonoscopies, imaging studies, lead tests, and prenatal encounters
Missed clinical statements for exclusion criteria	14	0	Reviewer missed pregnancy diagnosis, bilateral mastectomy procedure, 'absence of cervix' diagnosis, ischemic vascular disease diagnosis, myocardial infarction diagnosis, previous low back pain diagnosis, immunodeficiency diagnosis, and beta blocker contraindications
Miscounted clinical statements	8	0	Reviewer miscounted number of rotavirus vaccinations administered, outpatient prenatal visits, and well-care visits
Mismatch between claims and EHR data or medical coding errors	5	1	In cases where claims data did not match EHR data, we allowed EHR data to trump claims data only if it is explicitly allowed in the HEDIS specification. For example, we used the claims data for the following situations. According to the EHR, the patient had a nurse vaccination, but according to the data warehouse, it was an outpatient visit with Current Procedural Terminology (CPT) code 99212. According to the EHR, the patient did not have diabetes, but according to the data warehouse, the patient had a diabetes diagnosis with the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD9CM) code 250.00. According to the EHR, the patient had a family history of cancer, but according to the data warehouse the patient had a breast cancer diagnosis with ICD9CM code 174.9
Incorrectly assessed provider type	5	0	Reviewer thought that the provider was a non-relevant specialist, but the provider was also a general pediatrician, which is a suitable provider type for the CAP measure. Reviewer counted a visit which was performed by a provider type which was not allowed by the specification
Incorrect exclusion	3	1	Reviewer excluded a deceased patient, while it was not specified in the measure definition. Reviewer excluded a patient who had a hysterectomy performed after the evaluation date. CDC hemoglobin A1c measure requires excluding patients with chronic kidney disease stage 4; however, the reviewer excluded a patient with chronic kidney disease stage 3
Incorrect inclusion	4	0	Reviewer included a patient who was younger than the inclusion age limit. Reviewer included tests which were not in the allowed value set
Miscounted medication days' supply	4	0	MPM measure group requires counting number of days covered by at least one medication. Reviewer summed up treatment days for several drugs independently. Reviewer miscalculated number of days' supply
Incorrectly assessed medication	3	0	Reviewer counted some medications which were not in the allowed value set
Incorrectly assessed value	3	0	Reviewer missed latest blood pressure (BP) measurement, and used previous BP value instead
Incorrectly assessed date	2	0	COL measure requires a colonoscopy from less than 10 years ago; however, reviewer counted colonoscopy, which was performed 11 years ago. PPC measure evaluates whether there was at least one prenatal visit during the first trimester; however, reviewer counted a visit which occurred in the beginning of the second trimester
Incorrectly assessed clinical statements	1	0	In MPM measure group, reviewer incorrectly concluded that one of the encounters had the lab panel test performed
Quality measure specification allows for an unintended clinical scenario	1	0	In LBP measure, patient was diagnosed with a low back pain diagnosis and also had hip and pelvis X-ray procedure performed with Uniform Bill Revenue (UBREV) code 0320 which was incorrectly counted as a low back imaging study
Human errors when electronic phenotyping results were wrong (N = 18)			
Incorrect inclusion	0	7	The CBP measure logic was initially implemented incorrectly. After correcting the logic, we found that some patients were initially included without a hypertension diagnosis
Miscounted clinical statement	4	3	In the FPC and PPC measure groups, the electronic system initially miscounted visits. After correcting the error, we found out that the reviewer miscounted some visits as well
Incorrectly assessed value	0	2	After adding to the data warehouse additional body mass indexes, which were initially available only in the EHR, we found that the reviewer had initially missed some of these values
Missed clinical statement	1	1	Reviewer missed a hypertension diagnosis, and an external lead test

study at other institutions and for other phenotyping contexts is recommended to further evaluate the optimal strategy for single-reviewer phenotyping validation. Furthermore, this study was conducted by a single reviewer which might limit the generalizability of this study. However, (1) we feel that the reviewer selected is representative of a typical chart reviewer, (2) the main study insights are useful even if the magnitude of findings may differ for other reviewers, and (3) the methodology used could be of general utility for institutions to empirically optimize their single-reviewer auditing strategy. Yet another limitation is that neither the reviewer nor the adjudication team were blinded to the review strategy used. Finally, the reviewer did not have formal medical education. However, the reviewer had 16 years of experience conducting validations for billing, coding and quality measurement, and she was the designated expert for operational phenotyping validation within our institution's medical group.

6. Conclusion

Overall, our study shows that a single-reviewer validation process can be successfully used to reach and sustain high quality phenotyping in an operational setting. Our estimates suggest that providing electronic phenotyping results makes the chart review process more efficient and effective overall. On the other hand, the control strategy could allow finding more electronic phenotyping errors if the number of available cases to review is restricted. We recommend using a mix of phenotyping validation strategies, with the balance of strategies (i.e., ratio of control and intervention chart review strategies) based on the true electronic phenotyping error rate, the tolerance for missed electronic phenotyping errors, as well as the expertise, cost, and availability of personnel involved in chart review and discrepancy adjudication. Using a mix of review strategies leverages the strengths of both strategies.

Acknowledgments

KK is or has been a consultant or sponsored researcher on clinical decision support (CDS) or electronic clinical quality measurement to the U.S. Office of the National Coordinator for Health IT, ARUP Laboratories, McKesson InterQual, Hitachi, ESAC, Inc., JBS International, Inc., Inflexxion, Inc., Intelligent Automation, Inc., Partners HealthCare, Mayo Clinic, and the RAND Corporation. KK receives royalties for a Duke University-owned CDS technology for infectious disease management known as CustomID that he helped develop. KK was formerly a consultant for Religent, Inc. and a co-owner and consultant for Clinica Software, Inc., both of which provide commercial CDS services, including through use of a CDS technology known as SEBASTIAN that KK developed. KK no longer has a financial relationship with either Religent or Clinica Software. KK has no competing interest with any specific product or intervention evaluated in this manuscript. All other authors declare no competing interests.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] J. Pathak, A.N. Kho, J.C. Denny, Electronic health records-driven phenotyping: challenges, recent advances, and perspectives, J. Am. Med. Inf. Assoc. 20 (2013) e206–e211, http://dx.doi.org/10.1136/amiajnl-2013-002428.

[2] R.L. Richesson, W.E. Hammond, M. Nahm, D. Wixted, G.E. Simon, J.G. Robinson, A.E. Bauck, D. Cifelli, M.M. Smerek, J. Dickerson, R.L. Laws, R.A. Madigan, S.A. Rusincovitch, C. Kluchar, R.M. Califf, Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory, J. Am. Med. Inf. Assoc. 20 (2013) e226–e231, http://dx.doi.org/10.1136/amiajnl-2013-001926.

[3] R.L. Richesson, M. Smerek, Electronic health records-based phenotyping, NIH Heal. Care Syst. Res. Collab. (2014). <http://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/> (accessed July 22, 2015).

[4] C.-J. Hsiao, E. Hing, Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2013, NCHS Data Brief., 2014, pp. 1–8 <http://www.ncbi.nlm.nih.gov/pubmed/24439138> (accessed December 2, 2015).

[5] S.L. Cutrona, S. Toh, A. Iyer, S. Foy, E. Cavagnaro, S. Forrow, J.A. Racoosin, R. Goldberg, J.H. Gurwitz, Design for validation of acute myocardial infarction cases in Mini-Sentinel, Pharmacoepidemiol. Drug Saf. 21 (suppl. 1) (2012) 274–281, http://dx.doi.org/10.1002/pds.2314.

[6] J. Pathak, K.R. Bailey, C.E. Beebe, S. Bethard, D.C. Carrell, P.J. Chen, D. Dligach, C. M. Endle, L.A. Hart, P.J. Haug, S.M. Huff, V.C. Kaggal, D. Li, H. Liu, K. Marchant, J. Masanz, T. Miller, T.A. Oniki, M. Palmer, K.J. Peterson, S. Rea, G.K. Savova, C.R. Stancl, S. Sohn, H.R. Solbrig, D.B. Suesse, C. Tao, D.P. Taylor, L. Westberg, S. Wu, N. Zhuo, C.G. Chute, Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium, J. Am. Med. Inf. Assoc. 20 (2013) e341–e348, http://dx.doi.org/10.1136/amiajnl-2013-001939.

[7] D.A. Hanauer, Q. Mei, J. Law, R. Khanna, K. Zheng, Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE), J. Biomed. Inf. 55 (2015) 290–300, http://dx.doi.org/10.1016/j.jbi.2015.05.003.

[8] M. Reed, J. Huang, I. Graetz, R. Brand, J. Hsu, B. Fireman, M. Jaffe, Outpatient electronic health records and the clinical care and outcomes of patients with diabetes mellitus, Ann. Int. Med. 157 (2012) 482–489, http://dx.doi.org/10.7326/0003-4819-157-7-201210020-00004.

[9] Measure Authoring Tool 2012 Update |User Guide| Glossary, Natl. Qual. Forum., 2012 <http://C:/Users/Polina/Downloads/measure_authoring_tool_glossary-of-terms(1).pdf> (accessed June 2, 2016).

[10] C.W. Carspecken, P.J. Sharek, C. Longhurst, N.M. Pageler, A clinical case of electronic health record drug alert fatigue: consequences for patient outcome, Pediatrics 131 (2013) e1970–e1973, http://dx.doi.org/10.1542/peds.2012-3252.

[11] A. Parsons, C. McCullough, J. Wang, S. Shih, Validity of electronic health record-derived quality measurement for performance monitoring., J. Am. Med. Inform. Assoc. 19 604–609, doi:http://dx.doi.org/10.1136/amiajnl-2011-000557.

[12] L.P. Casalino, G.C. Alexander, L. Jin, R.T. Konetzka, General internists' views on pay-for-performance and public reporting of quality scores: a national survey, Health Aff. (Millwood). 26, 492–499, doi:http://dx.doi.org/10.1377/hlthaff.26.2.492.

[13] K.M. Newton, P.L. Peissig, A.N. Kho, S.J. Bielinski, R.L. Berg, V. Choudhary, M. Basford, C.G. Chute, I.J. Kullo, R. Li, J.A. Pacheco, L.V. Rasmussen, L. Spangler, J.C. Denny, Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network, J. Am. Med. Inf. Assoc. 20 (2013) e147–e154, http://dx.doi.org/10.1136/amiajnl-2012-000896.

[14] L.M. Kern, S. Malhotra, Y. Barrón, J. Quaresimo, R. Dhopeshwarkar, M. Pichardo, A.M. Edwards, R. Kaushal, Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study, Ann. Int. Med. 158 (2013) 77–83, http://dx.doi.org/10.7326/0003-4819-158-2-201301150-00001.

[15] A. Amster, J. Jentzsch, H. Pasupuleti, K.G. Subramanian, Completeness, accuracy, and computability of National Quality Forum-specified eMeasures, J. Am. Med. Inf. Assoc. 22 (2015) 409–416, http://dx.doi.org/10.1136/amiajnl-2014-002865.

[16] M.H. Stanfill, M. Williams, S.H. Fenton, R.A. Jenders, W.R. Hersh, A systematic literature review of automated clinical coding and classification systems., J. Am. Med. Inf. Assoc. 17, 646–651, doi:http://dx.doi.org/10.1136/jamia.2009.001024.

[17] J.A. Linder, E.O. Kaleba, K.S. Kmetik, Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set, Med. Care. 47 (2009) 208–216, http://dx.doi.org/10.1097/MLR.0b013e318189375f.

[18] K.E. Walsh, S.L. Cutrona, S. Foy, M.A. Baker, S. Forrow, A. Shoaibi, P.A. Pawloski, M. Conroy, A.M. Fine, L.E. Nigrovic, N. Selvam, M.S. Selvan, W.O. Cooper, S. Andrade, Validation of anaphylaxis in the Food and Drug Administration's Mini-Sentinel, Pharmacoepidemiol. Drug Saf. 22 (2013) 1205–1213, http://dx.doi.org/10.1002/pds.3505.

[19] S.H. Brown, P.L. Elkin, S.T. Rosenbloom, E. Fielstein, T. Speroff, eQuality for all: extending automated quality measurement of free text clinical narratives, in: AMIA Annual Symposium of Proceedings, 2008, pp. 71–75 <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2656015&tool=pmcentrez&rendertype=abstract> (accessed March 18, 2015).

[20] J.B. Shelton, T.A. Skolarus, D. Ordin, J. Malin, A. Antonio, J. Ryoo, C.S. Saigal, Validating electronic cancer quality measures at Veterans Health Administration, Am. J. Manag. Care. 20 (2014) 1041–1047. <http://www.ncbi.nlm.nih.gov/pubmed/25526392> (accessed January 6, 2015).

[21] T. Garrido, S. Kumar, J. Lekas, M. Lindberg, D. Kadiyala, A. Whippy, B. Crawford, J. Weissberg, E-Measures: insight into the challenges and opportunities of automating publicly reported quality measures, J. Am. Med. Inf. Assoc. 21 (2014) 181–184, http://dx.doi.org/10.1136/amiajnl-2013-001789.

[22] S.V.S. Pakhomov, J.D. Buntrock, C.G. Chute, Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques, J. Am. Med. Inf. Assoc. 13 (n.d.) 516–525, doi:http://dx.doi.org/10.1197/jamia.M2077.

[23] University of Utah Health Care - Salt Lake City, Utah, n.d. <http://healthcare.utah.edu/about/> (accessed January 30, 2014).

[24] P. Kukhareva, K. Kawamoto, D.E. Shields, D.T. Barfuss, A.M. Halley, T.J. Tippetts, P.B. Warner, B.E. Bray, C.J. Staes, Clinical decision support-based quality measurement (CDS-QM) framework: prototype implementation, evaluation, and future directions, in: AMIA Annual Symposium of Proceedings, 2014, pp. 825–834.

[25] K. Kawamoto, D. Shields, G. Del Fiol, OpenCDS: Enabling Clinical Decision Support at Scale through Open-Source, Standards-Based Software and Resources, AMIA Poster, 2011, p. 1830 <http://assets.conferencespot.org/fileserver/file/121836/filename/1lqa1c.pdf> (accessed February 24, 2014).

[26] NCQA, HEDIS 2015, 2014 <http://www.ncqa.org/hedis-quality-measurement/hedis-measures/hedis-2015>.

[27] The Groovy Programming Language, Off., n.d. <http://www.groovy-lang.org/> (accessed September 23, 2015).

[28] M. Chang, Monte Carlo Simulation for the Pharmaceutical Industry: Concepts, Algorithms, and Case Studies, CRC Press, 2011.

[29] C.J. McDonald, Protocol-based computer reminders, the quality of care and the non-perfectability of man, N. Engl. J. Med. 295 (1976) 1351–1355, http://dx.doi.org/10.1056/NEJM197612092952405.

[30] K. Kawamoto, C.A. Houlihan, E.A. Balas, D.F. Lobach, Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success, BMJ 330 (2005) 765, http://dx.doi.org/10.1136/bmj.38398.500764.8F.

CHAPTER 5

DISCUSSION

This dissertation consists of 3 interrelated research studies aimed at advancing computer-facilitated clinical quality improvement. Chapter 2 describes a qualitative study in which domain experts were interviewed to identify opportunities and challenges in advancing clinical quality improvement through the coordinated integration of CDS and eCQM. This study established the need for better integration of CDS and eCQM, identified benefits and challenges to integration of CDS and eCQM, and proposed approaches to addressing these challenges. Chapter 3 addresses one of the main challenges described in the first study – the lack of a standard-based framework that would allow implementation of CDS and eCQM in the same fashion.[1] A CDS framework called OpenCDS[2] was successfully used to support eCQM. However, a capability to implement both CDS and eCQM using the same framework did not guarantee high accuracy in the generated electronic phenotypes. Indeed, low accuracy of electronic phenotyping was one of the key problems identified by the domain experts in the first study. The last study in this dissertation, described in Chapter 4, investigated how to most effectively improve the accuracy of electronic phenotypes in operational settings.[3] Taken together, these studies have advanced the science of the use of informatics in healthcare.

5.1 Concurrent Efforts by Others

As a testament to the importance of this topic, several other groups were actively engaged in related efforts during the timeframe of this dissertation research. In particular, there were significant ongoing standards development and validation efforts in the areas of CDS, eCQM, and CDS-eCQM harmonization. These relevant standards development efforts are described below.

One of the most notable standards development efforts was the Clinical Quality Framework (CQF) initiative, a public-private partnership sponsored by the Centers for Medicare & Medicaid Services (CMS) and the Office of the National Coordinator for Health Information Technology (ONC) to identify, develop, and harmonize standards for CDS and eCQM.[4] The CQF work group developed and tested the HL7 Clinical Quality Language (CQL) standard to enable representing computable expression logic for both CDS and eCQM.[5] The Clinical Quality Language Specification, Release 1 was published in May 2015 as an HL7 Standard for Trial Use. CQF also worked on the Quality Improvement and Clinical Knowledge (QUICK) data model to represent patient data for CDS and eCQM, as well as a variety of standards based on the HL7 Fast Healthcare Interoperability Resources (FHIR) standard.[6] These FHIR-based standards include the FHIR Clinical Reasoning module and the FHIR QICore Implementation Guide.[7]

In another highly relevant initiative, the HL7 Clinical Information Modeling Initiative (CIMI) Work Group is developing detailed clinical models that can serve as the foundation of other standards, including FHIR profiles.[8] The HL7 CDS and Clinical Quality Improvement (CQI) Work Groups are working with the HL7 CIMI Work Group to enable a rigorous foundation of data interoperability to support CDS and eCQM.

5.2 Context Within Continuous Clinical Quality Improvement

CDS and eCQM are a component of the larger context of continuous quality improvement. According to the Institute of Medicine, healthcare organizations should transform into learning healthcare systems (LHS) through such continuous and systematic efforts to measure and improve care quality.[9] The Institute of Medicine suggests that the patient care experience should be systematically captured, assessed, and translated into reliable care. The LHS is based on accountability and feedback which allow virtuous cycles. Due to the "imperfectability of men,"[10] perfect healthcare cannot be achieved without relying on computers. Integration of CDS and eCQM and improved validation strategies can simplify the automation required to support a LHS.

5.3 Significance

This dissertation contributes significantly to the field of computer-facilitated clinical quality improvement. Advancing CDS and eCQM is essential to improving care quality and bending the cost curve. Integration of CDS and eCQM has the potential to improve medical care because it allows the closing of the feedback loop for the quality improvement cycles and simplifies the development, implementation, and maintenance of machine-executable knowledge for both CDS and eCQM. Reduced duplication of effort could help to enable greater progress in quality improvement in the face of limited available resources. Furthermore, CDS could help improve the accuracy of eCQMs by enabling the point-of-care collection of data points relevant for eCQMs, such as exclusion conditions for care interventions. In summary, a unified and validated CDS-QM framework could facilitate the provision of higher quality care within the larger

context of continuous quality improvement and the LHS.

5.4 Innovation

The work presented in this dissertation is innovative because it provides a new vision and a new framework for quality improvement in healthcare. Even though the number of publications about CDS is large, associated quality measurement efforts rarely use the same underlying technical approach.[11] To the best of our knowledge, there is only a limited number of papers in the peer-reviewed literature which describe the software architecture, implementation issues, and cultural challenges associated with simultaneous implementation of performance measures and corresponding CDS interventions in a broad spectrum of healthcare related organizations.[12,13] Moreover, existing manuscripts describe experiences within specific organization which may not be directly generalizable,[12,13] whereas our qualitative study interviewed domain experts from numerous organizations to gather more generalizable insights. The double independent human expert review approach, with adjudication performed for interreviewer discrepancies, is generally considered the gold standard for electronic phenotyping validation in research settings.[14–16] However, such double review is generally not feasible in operational settings, and we overcame this challenge by proposing and validating an innovative pragmatic single reviewer validation framework which could be used in routine operational settings. Finally, while there were a handful of prior studies that used the same underlying technology for both CDS and eCQM,[12,17,18] we were one of the only ones to accomplish this integration using a standards-based, open-source approach.

5.5 Limitations

This research has some limitations. First, we were unable to address all the challenges in computer-facilitated quality improvement. However, the field is so immense that no single body of work can adequately address all the current challenges. Second, Chapters 2 and 3 are based on research carried out in a single academic hospital. However, University of Utah Health Care is representative of many other academic hospitals and we believe that the study findings should be generalizable to other care settings.

5.6 Future Directions

There are many outstanding issues remaining for improving care through CDS and eCQM, and the recommendations synthesized from domain experts could be used to guide future work. In particular, the integration of CDS and eCQM is still in its early stages, requiring significant continued work to impact care broadly. In particular, as was noted by the domain experts in Chapter 2, there is still significant heterogeneity in data representation across health IT systems and healthcare institutions. Such heterogeneity must be addressed if CDS and eCQM are to be truly interoperable. Currently, the most promising approach for addressing this long-standing issue appears to be the use of detailed FHIR profiles based on CIMI models, so that a widely adopted data interoperability approach (FHIR) can be coupled with the level of detailed semantics required for true interoperability. While the definition of such detailed FHIR profiles and underlying CIMI models still will not fully address issues of different clinical workflows and associated data collection methodologies, as well as differences in data already

collected in different means (if they cannot be mapped 1:1 to these detailed models), the first step must be the definition of such detailed models.

With regard to the CDS-eCQM framework, a natural progression would be to update the data model from the vMR to FHIR. Also, the CDS service framework could be updated to use the CDS Hooks[19] specification rather than the Decision Support Service specification, given the increasing adoption of CDS Hooks by EHR vendors. Indeed, active efforts are currently underway at the University of Utah to make this transition in the CDS-eCQM framework.

In the area of electronic phenotype validations, a potential future direction is to develop cross-institutional applications for enabling electronic validation of phenotypes in operational settings. Underlying these validations will need to be accurate phenotyping that can be scaled, which potentially could be accomplished through the use of detailed FHIR profiles as well as scalable CDS-eCQM evaluation approaches as described in Chapter 3. Using these phenotyping results, a Substitutable Medical Applications and Reusable Technologies (SMART) application could be developed for enabling a validation framework fully integrated with the EHR, thereby facilitating the necessary human chart reviews.[20,21]

In addition, moving forward, the work presented in this dissertation should be validated in other institutions to ensure generalizability and broad applicability. Once validated, the hope would be that this work will be able to influence care widely across various healthcare settings.

5.7 References

1. Kukhareva P V, Kawamoto K, Shields DE, et al. Clinical decision support-based quality measurement (CDS-QM) framework: prototype implementation, evaluation, and future directions. *AMIA Annu Symp Proc*. 2014;2014:825-834.

2. OpenCDS Home. http://www.opencds.org/. Accessed January 30, 2014.

3. Kukhareva P, Staes C, Noonan KW, et al. Single-reviewer electronic phenotyping validation in operational settings: Comparison of strategies and recommendations. *J Biomed Inform*. 2017;66:1-10. doi:10.1016/j.jbi.2016.12.004.

4. Kawamoto K, Hadley MJ, Oniki T, Skapik J. The clinical quality framework initiative to harmonize decision support and quality measurement standards: defined standards, pilot results, and moving beyond quality improvement. *AMIA Annu Symp Proc*. 2015:202-204.

5. HL7 Standards Product Brief - HL7 Cross-Paradigm Specification: Clinical Quality Language, Release 1. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=400. Accessed July 25, 2017.

6. HL7 Standards Product Brief - HL7 Fast Healthcare Interoperability Resources Specification (FHIR®), DSTU Release 2. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=448. Accessed July 25, 2017.

7. Quality Improvement Core (QI-Core) Implementation Guide - FHIR v1.6.0. http://hl7.org/FHIR/us/qicore/2016Sep/index.html. Accessed July 25, 2017.

8. Goossen WTF. Detailed clinical models: representing knowledge, data and semantics in healthcare information technology. *Healthc Inform Res*. 2014;20(3):163-172. doi:10.4258/hir.2014.20.3.163.

9. Institute of Medicine. *Leadership Commitments to Improve Value in Healthcare: Finding Common Ground*. Washington (DC): The National Academies Press; 2009.

10. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. *N Engl J Med*. 1976;295(24):1351-1355. doi:10.1056/NEJM197612092952405.

11. Lobach D, Sanders GD, Bright TJ, et al. Enabling health care decisionmaking through clinical decision support and knowledge management. *Evid Rep Technol Assess (Full Rep)*. 2012;(203):1-784.

12. Goldstein MK, Tu SW, Martins S, et al. Automating performance measures and clinical practice guidelines: differences and complementarities. *AMIA Annu Symp Proc*. 2014:37-38.

13. Haggstrom DA, Saleem JJ, Militello LG, Arbuckle N, Flanagan M, Doebbeling BN. Examining the relationship between clinical decision support and performance measurement. *AMIA Annu Symp Proc*. 2009;2009:223-227.

14. Richesson RL, Smerek M. Electronic health records-based phenotyping. NIH Health Care Systems Research Collaboratory. http://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/. Published 2014. Accessed July 22, 2015.

15. Brown SH, Elkin PL, Rosenbloom ST, Fielstein E, Speroff T. eQuality for all: extending automated quality measurement of free text clinical narratives. *AMIA Annu Symp Proc*. January 2008:71-75.

16. Linder JA, Kaleba EO, Kmetik KS. Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. *Med Care*. 2009;47(2):208-216. doi:10.1097/MLR.0b013e318189375f.

17. Raja AS, Gupta A, Ip IK, Mills AM, Khorasani R. The use of decision support to measure documented adherence to a national imaging quality measure. *Acad Radiol*. 2014;21(3):378-383. doi:10.1016/j.acra.2013.10.017.

18. Lakshminarayan K, Rostambeigi N, Fuller CC, Peacock JM, Tsai AW. Impact of an electronic medical record-based clinical decision support tool for dysphagia screening on care quality. *Stroke*. 2012;43(12):3399-3401. doi:10.1161/STROKEAHA.112.662536.

19. CDS Hooks. http://cds-hooks.org/. Accessed September 14, 2017.

20. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *J Am Med Informatics Assoc*. 2016;23(5):1-10. doi:10.1093/jamia/ocv189.

21. SMART Health IT – Connecting health system data to innovators' apps. https://smarthealthit.org/. Accessed September 14, 2017.

CHAPTER 6

CONCLUSION

The overarching goal of this research was to advance computer-facilitated clinical quality improvement. Within this larger goal, this work aimed to address the lack of integration of CDS and eCQM and the inadequate accuracy of electronic phenotyping. The aims of this dissertation were achieved by (1) conducting a qualitative study of domain experts which explored beliefs and perceptions regarding the integration of CDS and eCQM functionality and activities, (2) demonstrating the feasibility of implementing eCQM using a CDS infrastructure, and (3) evaluating pragmatic strategies for single human validation of electronic phenotype evaluation results in operational settings.

This research succeeded in exploratory analysis of issues related to CDS-eCQM integration; proposed and evaluated a standard-based, open-source CDS-eCQM framework; and evaluated 2 approaches to single-reviewer validation of electronic phenotyping results. This dissertation represents a significant step towards understanding and addressing barriers to the integration and validation of CDS and eCQM.

Computer-facilitated quality improvement is an active, growing, and constantly changing field. While many challenges remain in the use of computer-facilitated quality improvement, this dissertation suggests solutions and approaches that could be followed

to improve the quality of healthcare using informatics. It is hoped that results from this dissertation, along with other projects currently ongoing in this field, including FHIR and CIMI, will inform new strategies for enhancing the efficiency and accuracy of computer-facilitated quality improvement, thereby ultimately leading to improvements in care quality in the United States and beyond.