

Adapting SQUIRE 2.0 to Create a Quality Improvement Evidence-Based Medicine Critical Appraisal Tool (QI-EBM-CAT) for Graduate Medical Education Trainees

Andrea Smeraglio¹, MD
 Brook Pittenger, MD
 Matthew DiVeronica, MD
 Bryn McGhee, MD

Christopher Terndrup, MD
 Ramya J. Prasad, DO
 Patricia A. Carney, PhD, MS
 Greg Ogrinc, MD, MS

ABSTRACT

Background Evidence-based medicine (EBM) has long been taught to physician trainees for critical appraisal of research manuscripts. There is no parallel or similar framework to guide trainees in the appraisal of quality improvement (QI) literature.

Objective To adapt existing guidelines of QI manuscript reporting into an educational QI-EBM appraisal tool to help residents distinguish research and QI manuscripts, assess QI designs and methodologies, and evaluate QI manuscripts' strengths and weaknesses.

Methods Between 2018 and 2021, we developed a QI-EBM critical appraisal tool (QI-EBM-CAT) and performed 3 plan-do-study-act cycles to refine the tool based on JAMA and SQUIRE 2.0 guidelines. We then surveyed residents regarding the usefulness of the tool and their confidence in evaluating QI manuscripts before and after completing a QI-EBM workshop using the QI appraisal tool.

Results Sixty-six of 74 internal medicine postgraduate year (PGY)-1 to PGY-3 residents (89.2%) completed the workshop and assessment surveys in 2021. The workshop was found to be moderately to very useful by 85.1% (63 of 74) of residents as a framework for QI manuscript critical analysis. The summary confidence score in QI manuscript critical appraisal improved from a 64% rating of moderately to very confident in the pre-period to 94.6% in the post-period ($P < .001$) with statistical improvements in all 5 confidence areas assessed ($P < .001$).

Conclusions The QI-EBM-CAT, designed to teach residents how to critically assess QI manuscripts using EBM principles, resulted in subjective improvements in confidence of QI manuscript analysis.

Introduction

Evidence-based medicine (EBM), broadly defined as the judicious integration of current best evidence into the delivery of medical care, is a common component of graduate medical education (GME) curricula.¹ The Accreditation Council for Graduate Medical Education (ACGME) endorses Practice-Based Learning and Improvement as 1 of 6 core competencies in their requirements for GME programs; teaching EBM practices to residents is an assessment milestone within this domain.² Critical appraisal of clinical research literature, including randomized controlled trials (RCTs), meta-analyses, and systematic reviews, is a central skill in EBM and is routinely taught in the context of a traditional journal club, guided by well-defined best practices and instructional frameworks.³⁻⁵ In contrast, teaching critical appraisal of

quality improvement (QI) literature to residents lacks similarly robust approaches and tools to inform instructional design and is further complicated by the incompatibility of applying clinical research literature appraisal methods to a QI manuscript.⁶⁻⁹ While tools for QI manuscript development have been established, most notably SQUIRE 2.0,¹⁰ these have yet to be translated into an educational context.

At our institution, internal medicine residents engage in a longitudinal EBM curriculum in a journal club setting that focuses on clinical research literature appraisal and developing skills to interpret therapeutic and diagnostic evidence. In 2018, our QI faculty were asked by EBM curricular leads to develop an approach that integrates QI literature appraisal into the EBM curriculum. In the planning process, we identified a notable gap in the ability of residents to interpret QI literature during our residents' concurrent QI course. Faculty noted that residents did not understand the design and methodology differences between QI and clinical research papers and lacked skills to assess a QI paper's strengths and limitations.

DOI: <http://dx.doi.org/10.4300/JGME-D-22-00033.1>

Editor's Note: The online version of this article contains the Quality Improvement Evidence-Based Medicine Critical Appraisal Tool and survey used in the study.

Given these findings, as well as the ACGME's shift toward preparation of graduates as systems thinkers,¹¹ we created an educational intervention for residents comprised of a systematic QI literature assessment tool. The intervention was designed to achieve several fundamental objectives in teaching residents: (1) Distinguish the differing approaches to QI and clinical research literature appraisal; (2) Teach basic QI study design, methodology, and interpretation skills; and (3) Provide a framework for assessing the strengths and weaknesses of QI interventions. Here, we report on the iterative development of a QI-EBM critical appraisal tool (QI-EBM-CAT), how residents rated the usefulness of the tool, and changes in resident confidence in appraising QI literature.

Methods

Study Setting

This study was conducted within the internal medicine residency program (~100 residents) at Oregon Health & Science University (OHSU), a metropolitan academic health center located in the Pacific Northwest, with approximately 100 residents. Residents at OHSU participate in a year-long, 150-minute, monthly, longitudinal, health systems science curriculum through which the basics of QI methodology are taught in addition to other systems thinking skills. Thus, all residents involved in this study, except interns who had not yet completed the curriculum, had basic QI knowledge at the start of the study.

The residency program has an additional required, monthly, 90-minute, research EBM curriculum that exists separate from the health systems science course. It was into the EBM curriculum that the QI-EBM-CAT was introduced. The QI-EBM-CAT replaced 1 to 2 sessions annually of the research EBM curriculum from 2018 to 2020 (2018-2019 in-person; 2020 virtual format). The sessions included 20 to 30 participants, a mixture of postgraduate year (PGY)-1 to PGY-3 residents, and were taught over 4 consecutive weeks to capture all learners in the program during their respective weeks of continuity clinic. The QI-EBM-CAT and a single QI article were emailed to the residents 1 week prior to the session. From 2018 to 2020 each session featured a different QI article chosen by the Chief Resident in Quality and Patient Safety (CRQS) from the affiliated Portland Veterans Affairs (VA). The QI-EBM-CAT tool was then introduced to and used by residents to actively guide their appraisal of the selected QI paper.

Four faculty, consisting of 3 hospitalists, a primary care physician, and the VA CRQS, served as facilitators for all QI-EBM-CAT sessions. All facilitators also teach the longitudinal health systems science

Objectives

To adapt existing guidelines of quality improvement (QI) manuscript reporting into an educational evidence-based medicine (EBM) appraisal tool to train residents to critically analyze QI manuscripts.

Findings

We developed a QI-EBM critical appraisal tool (QI-EBM-CAT) that was able to improve the summary confidence scores improved from 64% pre- to 94.6% post-period ($P < .001$) of trainees reporting they felt moderately to very confident in QI manuscript critical appraisal.

Limitations

This is a single-site study that assessed subjective resident perception.

Bottom Line

The QI-EBM-CAT, designed to teach residents how to assess QI manuscripts using EBM principles, resulted in subjective improvements in confidence of QI manuscript analysis and aligns with the ACGME shift toward preparing residents in systems thinking.

curriculum and are experienced instructors at the intersection of QI and medical education. No additional funding was provided for faculty to develop or teach this course.

Intervention

We conducted 3 plan-do-study-act (PDSA) cycles to refine the QI-EBM-CAT (TABLE 1).

During PDSA cycle 1 (2018), the initial iteration of the QI-EBM-CAT was designed to provide organizational structure for residents to easily use during the QI-EBM teaching session. We adapted a list of key questions for appraising QI studies from the *Journal of the American Medical Association* (JAMA) User's Guide to the Medical Literature.¹² Selected questions were then arranged under the broad headings of Background, Validity, and Results to create the basic structure of the tool. Blank space was provided for note-taking as they read the manuscript (the final QI-EBM-CAT tool is provided as online supplementary data). The session was divided into 3 parts: 20 minutes for article review, 30 minutes for small group appraisal (each group assigned one of Background, Validity, or Results) and completion of the QI-EBM-CAT, and 20 to 30 minutes for large group discussion of each group's answers with faculty facilitation of the debrief.

During PDSA cycle 2 in 2019, subjective faculty evaluations of the QI-EBM teaching sessions and informal participant feedback from PDSA cycle 1 were compiled, and they both suggested the tool did not fully meet the aim of residents being able to discern between clinical research and QI literature. Thus, we added a 3- to 5-minute introductory didactic called the QI-EBM primer, in which faculty

TABLE 1

QI-EBM-CAT Development and Training Timeline

PDSA Cycle	QI-EBM-CAT Changes	Results of PDSA Cycle
PDSA 1: 2018	- QI-EBM-CAT was adapted from JAMA User's Guide to the Medical Literature - Structure included: Background, validity and bias, results and applications - In-person teaching sessions - Team-based learning	Faculty feedback suggested the tool did not adequately assist residents in distinguishing between clinical research and QI papers, prompting addition of QI-EBM primer
PDSA 2: 2019	- PDSA 1 components preserved - Added QI-EBM primer didactic	Residents lacked adequate understanding and interpretation of methodology section of QI papers, prompting restructuring of QI-EBM-CAT tool based on SQUIRE 2.0 guidelines
PDSA 3: 2020	- QI-EBM-CAT revised with adaption of SQUIRE 2.0 - Revised structure included: introduction, methods, results, discussion - Subheadings added - Transition to virtual teaching sessions in setting of COVID-19 pandemic	Faculty and resident feedback were positive about utility of revised QI-EBM-CAT

Abbreviations: QI-EBM-CAT, Quality Improvement Evidence-Based Medicine Critical Appraisal Tool; PDSA, plan-do-study-act; JAMA, *Journal of the American Medical Association*; SQUIRE, Standards for Quality Improvement Reporting Excellence.

described salient structural and interpretive differences between clinical research and QI papers (TABLE 2).¹³ The remainder of the intervention remained unchanged from PDSA cycle 1.

During PDSA cycle 3 (2020), we revised the QI-EBM-CAT to enable residents to translate QI-EBM appraisal knowledge and skills into future QI project

participation and publication. We used the revised Standards for Quality Improvement Reporting Excellence (SQUIRE 2.0) to rework the guiding questions and further emphasize understanding methodology.¹⁰ The SQUIRE 2.0 guidelines were closely followed, except for educational additions to highlight the differences between QI and research manuscripts, and

TABLE 2

QI-EBM Primer: How Appraisal of Medical Research and QI Intervention Articles Differ

Article Section	Clinical Research	QI Intervention
Introduction	- Asks a clinical question - Previous literature that indicates the need for the study	- Frames the need for an intervention by defining a problem - Motivation for change - What literature exists on this problem? What interventions have been studied?
Methods	- Typically uses randomized, controlled designs - Multi-site study	- Usually non-randomized, non-controlled - Typically uses pre-post repeated measures design - Single-site study
Results	- Enumerative data - Static conditions → random samples can be used to make estimates of the group - The group is being studied - Analysis tools: chi-square, <i>t</i> test, ANOVA, correlational, multivariable regression	- Analytic data - Changing conditions → a random sample may not reflect the group due to changes over time - The process is being studied - Analysis tools: descriptive statistics, run charts, control charts
Discussion	- Was the primary outcome relevant to the clinical question? - Strengths of the study - Limitations of the study - Risk of bias	- Was the primary outcome patient-centered? - Replicable at other sites? - Sustainable?

Abbreviations: QI, quality improvement; EBM, evidence-based medicine; ANOVA, analysis of variance.

minor deletions of subsections for the sake of time efficiency. The QI-EBM-CAT deviated most notably from SQUIRE 2.0 in the results section, with additions to help trainees distinguish important differences from research papers, including how results are presented, strengths and weaknesses of presentation methods, and follow-up to ensure change is maintained across time. SQUIRE 2.0 language was modified into directed questions that encourage residents to seek specific answers from the manuscript to foster better understanding. Structural headings for the QI-EBM-CAT were rearranged into Introduction, Methods, Results, and Discussion, and subheadings were added to frame the rationale underlying each question. During this cycle, residents were divided into 4 small groups rather than 3, but the remainder of the session was unchanged from PDSA cycles 1 and 2.

Survey Development

We designed a 6-item survey to assess usefulness (1 item) of the QI-EBM-CAT session and confidence (5 items) in evaluating a QI manuscript before and after completing the QI-EBM workshop. A Likert-type scale was used to assess participants' responses. Face validity was established after several revisions, based on expert review. The survey was then tested on chief residents who had previously taken the QI-EBM course to assess question clarity and order. The survey is available as online supplementary data.

Data Analyses

Descriptive statistics, including frequencies and percentiles, were calculated for each survey variable. An overall confidence score was calculated by summing the numbers in each response category for each variable and calculating the percent based on the total in the pre- vs post-session survey. This overall score reflects a summary of confidence in the study population. Differences in survey findings according to PGY level were assessed using chi-square tests. The Wilcoxon signed rank test was used to compare participants' pre- and post-session responses due to the non-normal distribution of data. All tests were 2-tailed with the alpha level set at $<.05$ for determining statistical significance. IBM SPSS Statistics, version 28 was used to conduct analyses. OHSU's Institutional Review Board reviewed all study activities and considered them exempt due to their QI focus.

Results

Sixty-six of 74 participants (89.2%) responded to the survey and reported their training year and were

included in an unlinked analysis. Of these, 21 (31.8%) were PGY-1, 23 (34.8%) were PGY-2, and 22 (33.3%) were PGY-3. The majority of participants (85.1%) rated the usefulness of the session for providing a framework to appraise QI studies as somewhat to tremendously useful—somewhat useful (28.4%), quite a bit useful (43.2%), or tremendously useful (13.5%). No statistical differences were found between training years (P values $>.50$).

Overall changes in confidence for the 5 variables assessed between pre- and post-session are shown in TABLE 3. The summary confidence score in QI manuscript critical appraisal improved from a 64% rating of moderately to very confident in the pre-period to 94.6% in the post period ($P<.001$). In the pre-session period, the majority of participants rated themselves as being slightly to moderately confident (slightly confident range of 27.3%-34.8%; moderately confident range of 33.3%-42.4%). In the post-session period, the majority of participants reported being very confident (range of 47.5%-57.6%) and the P values for all 5 variables were $<.001$, indicating all increases in confidence were statistically significant. Again, no statistical differences were found between training years (P values for all 5 variables were $>.60$).

Discussion

To our knowledge, this is the first study to adapt existing QI literature guidelines into an educational QI-EBM critical appraisal tool to help trainees analyze QI manuscripts. Our QI-EBM session using the QI-EBM-CAT was perceived as moderately to very useful by trainees and resulted in a statistical improvement in all confidence items assessed as well as in the summary score for confidence related to analyzing a QI manuscript. ACGME's expanding guidelines around teaching QI to trainees and continued emphasis on utilizing evidence-based medicine makes these findings important.

The QI-EBM-CAT fills an important gap in the QI appraisal literature, as other contemporary QI appraisal tools do not focus on physician training. For example, the Quality Improvement Minimum Quality Criteria Set (QI-MQCS) provides guidelines for reviewers rather than learners.¹⁴ SQUIRE 2.0 is designed to provide guidelines for authorship.¹⁵ The QI-EBM-CAT is specifically a teaching tool in QI manuscript assessment and a framework for identifying strengths, weaknesses, methodology, and the differences between QI and research.

Our pre-course survey showed no differences in confidence scores when evaluating a QI paper across PGYs, suggesting that residents do not learn how to

TABLE 3
Changes in Confidence According to Session Component Between Pre-Session and Post-Session (n=66)^a

Survey Variable: How Confident Do You Feel When...	Pre-QI-EBM Session					Post-QI-EBM Session					P Value
	Not at All, n (%)	Slightly, n (%)	Moderately, n (%)	Very, n (%)	Extremely, n (%)	Not at All, n (%)	Slightly, n (%)	Moderately, n (%)	Very, n (%)	Extremely, n (%)	
Evaluating a QI manuscript	3 (4.5)	23 (34.8)	22 (33.3)	15 (22.7)	3 (4.5)	1 (1.5)	2 (3.0)	18 (27.3)	38 (57.6)	7 (10.6)	<.001
Identifying strengths in a QI manuscript	3 (4.5)	18 (27.3)	23 (34.8)	19 (28.8)	3 (4.5)	0 (0)	4 (6.1)	18 (27.3)	34 (51.5)	10 (15.2)	<.001
Identifying limitations in a QI manuscript	4 (6.1)	18 (27.3)	22 (33.3)	18 (27.3)	4 (6.1)	0 (0)	3 (4.5)	18 (27.3)	34 (51.5)	11 (16.7)	<.001
Recognizing the alignment between the question posed and methods used in a QI manuscript	4 (6.1)	18 (27.3)	28 (42.4)	12 (18.2)	4 (6.1)	0 (0)	4 (6.1)	17 (25.8)	38 (57.6)	7 (10.6)	<.001
Understanding the difference between a randomized controlled trial and a QI manuscript	4 (6.1)	14 (31.2)	28 (42.4)	14 (21.2)	6 (9.1)	1 (1.5)	3 (4.5)	12 (18.2)	38 (47.5)	12 (18.2)	<.001
Overall confidence ^b	18 (5.2)	105 (30.5)	123 (35.5)	78 (22.7)	20 (5.8)	2 (<1.0)	16 (4.9)	83 (25.2)	182 (55.2)	47 (14.2)	<.001

Abbreviations: QI, quality improvement; EBM, evidence-based medicine.

^a No statistical differences according to training year ($P > .60$).

^b Overall score represents the sum of responses in each response category with the percentage reflecting the distribution of scores based on a total of 100%.

evaluate QI papers from their traditional EBM education or basic QI education, highlighting the importance of this tool. Our survey suggests residents require specific training in QI-EBM appraisal to improve their confidence. Improving confidence has been linked to improved self-efficacy and ultimately authentic performance. When viewed through the lens of self-determination theory,¹⁶ the improved confidence our residents demonstrated after completing the QI-EBM course is an important first step of actualizing learned behaviors. Future research should focus on how training residents to interpret QI literature changes attitudes toward QI projects and involvement in QI initiatives. Although we only studied the utility and impact of this tool for trainees, the use of this tool could be generalizable to other medical education settings (such as undergraduate and continuing medical education) and should also be studied in those settings.

The limitations of this study include completion of the study at a single site with only internal medicine residents. In addition, the survey assessed residents' perceptions, a subjective measure. It should be noted that grouping the top 3 options on the 5-point Likert scale during our analysis may skew results positively. We also conducted an unlinked analysis; it would have been optimal to code the pre-post surveys so we could have linked survey responses in the pre- and post-periods to the same resident. Our sessions were taught by clinical educators with QI expertise, which could create context-specific advantages. Finally, the n dropped by 10.8% from pre- to post-assessment, which could affect survey outcomes.

Conclusions

We adapted SQUIRE 2.0 guidelines to create the first QI-specific EBM critical appraisal tool to train residents to critically assess a QI paper. We found statistically significant improvement in confidence with QI manuscript analysis among internal medicine residents.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-72. doi:10.1136/bmj.312.7023.71
2. Accreditation Council for Graduate Medical Education. Milestones Resources. Accessed June 10, 2021. <https://acgme.org/What-We-Do/Accreditation/Milestones/Resources/>
3. Lee AG, Boldt HC, Golnik KC, et al. Using the journal club to teach and assess competence in practice-based

- learning and improvement: a literature review and recommendation for implementation. *Surv Ophthalmol*. 2005;50(6):542-548. doi:10.1016/j.survophthal.2005.08.002
4. Al Achkar M. Redesigning journal club in residency. *Adv Med Educ Pract*. 2016;7:317-320. doi:10.2147/AMEP.S107807
 5. Schwartz MD, Dowell D, Aperi J, Kalet AL. Improving journal club presentations, or, I can present that paper in under 10 minutes. *Evid Based Med*. 2007;12(3):66-68. doi:10.1136/ebm.12.3.66-a
 6. Wong BM, Etchells EE, Kuper A, Levinson W, Shojania KG. Teaching quality improvement and patient safety to trainees: a systematic review. *Acad Med*. 2010;85(9):1425-1439. doi:10.1097/ACM.0b013e3181e2d0c6
 7. Mondoux S, Shojania KG. Evidence-based medicine: a cornerstone for clinical care but not for quality improvement. *J Eval Clin Pract*. 2019;25(3):363-368. doi:10.1111/jep.13135
 8. Massagli TL, Zumsteg JM, Osorio MB. Quality improvement education in residency training: a review. *Am J Phys Med Rehabil*. 2018;97(9):673-678. doi:10.1097/PHM.0000000000000947
 9. Brown A, Lafreniere K, Freedman D, et al. A realist synthesis of quality improvement curricula in undergraduate and postgraduate medical education: what works, for whom, and in what contexts? *BMJ Qual Saf*. 2021;30(4):337-352. doi:10.1136/bmjqs-2020-010887
 10. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*. 2016;25(12):986-992. doi:10.1136/bmjqs-2015-004411
 11. Lane-Fall MB, Davis JJ, Clapp JT, Myers JS, Riesenberg LA. What every graduating resident needs to know about quality improvement and patient safety: a content analysis of 26 sets of ACGME milestones. *Acad Med*. 2018;93(6):904-910. doi: 10.1097/ACM.0000000000002039.
 12. Fan E, Laupacis A, Pronovost PJ, Guyatt GH, Needham DM. How to use an article about quality improvement. *JAMA*. 2010;304(20):2279-2287. doi:10.1001/jama.2010.1692
 13. Al-Surimi K. Research versus quality improvement in healthcare. *Glob J Qual Saf Healthcare*. 2018;1(2):25-27. doi:10.4103/JQSH.JQSH_16_18
 14. Hempel S, Shekelle PG, Liu JL, et al. Development of the Quality Improvement Minimum Quality Criteria Set (QI-MQCS): a tool for critical appraisal of quality improvement intervention publications. *BMJ Qual Saf*. 2015;24(12):796-804. doi:10.1136/bmjqs-2014-003151
 15. McQuillan R, Wong B. The SQUIRE guidelines: a scholarly approach to quality improvement. *J Grad Med Educ*. 2016;8(5):771-772. doi:10.4300/JGME-D-16-00558.1
 16. Bandura A. Self-efficacy mechanism in human agency. *Am Psychol* 1982;37(2):122-147. doi:10.1037/0003-066X.37.2.122



Andrea Smeraglio, MD, is Assistant Professor of Internal Medicine, Oregon Health & Science University, School of Medicine; **Brook Pittenger, MD**, is Assistant Professor of Internal Medicine, University of Washington, School of Medicine; **Matthew DiVeronica, MD**, is Assistant Professor of Internal Medicine, Oregon Health & Science University, School of Medicine; **Bryn McGhee, MD**, is Assistant Professor of Internal Medicine, Oregon Health & Science University, School of Medicine; **Christopher Terndrup, MD**, is Assistant Professor of Internal Medicine, Oregon Health & Science University, School of Medicine; **Ramya J. Prasad, DO**, is PGY-4 Chief Resident in Quality & Safety, Oregon Health & Science University, School of Medicine; **Patricia A. Carney, PhD, MS**, is Professor of Family Medicine, Oregon Health & Science University, School of Medicine; and **Greg Ogrinc, MD, MS**, is Senior Vice President, Certification Standards and Programs, American Board of Medical Specialties, and Visiting Clinical Professor of Medicine, University of Illinois Chicago College of Medicine.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

The authors gratefully acknowledge Hayden Oldham for contributions to the QI-EBM-CAT creation.

Corresponding author: Andrea Smeraglio, MD, Oregon Health & Science University, smeragli@ohsu.edu, Twitter @anneghena

Received January 10, 2022; revision received May 2, 2022; accepted October 4, 2022.