Identifying and resolving the frustrations of reviewing the improvement literature: The experiences of two improvement researchers

Emma Jones, [1,2] Joy Furnival,[3,4] Wendy Carter[5]

[1]Clinical Trials Unit, University of Warwick, Coventry, UK
[2]Orthopaedic directorate, University Hospitals of Coventry and Warwickshire (NHS Trust), Coventry, United Kingdom
[3]Improvement Directorate, NHS Improvement, Waterloo House, London, UK
[4]Health Management Group, Alliance Manchester Business School, University of Manchester, Manchester, UK
[5]Maternity Services, Homerton University Hospital NHS Foundation Trust, London, UK

Correspondence to
Dr Emma Jones, clinical trials unit, University of Warwick, Coventry CV4 7AL, UK; e.jones.11@warwick.ac.uk

ABSTRACT

Background and aims Summarising quality improvement (QI) research through systematic literature review has great potential to improve patient care. However, heterogeneous terminology, poor definition of QI concepts and overlap with other scientific fields can make it hard to identify and extract data from relevant literature. This report examines the compromises and pragmatic decisions that undertaking literature review in the field of QI requires and the authors propose recommendations for literature review authors in similar fields.

Methods Two authors (EJ and JF) provide a reflective account of their experiences of conducting a systematic literature review in the field of QI. They draw on wider literature to justify the decisions they made and propose recommendations to improve the literature review process. A third collaborator, (WC) co-created the paper challenging author's EJ and JF views and perceptions of the problems and solutions of conducting a review of literature in QI.

Results Two main challenges were identified when conducting a review in QI. These were defining QI and selecting QI studies. Strategies to overcome these problems include: select a multi-disciplinary authorship team; review the literature to identify published QI search strategies, QI definitions and QI taxonomies; Contact experts in related fields to clarify whether a paper meets inclusion criteria; keep a reflective account of decision making; submit the protocol to a peer reviewed journal for publication.

Conclusions The QI community should work together as a whole to create a scientific field with a shared vision of QI to enable accurate identification of QI literature. Our recommendations could be helpful for systematic reviewers wishing to evaluate complex interventions in both QI and related fields.

INTRODUCTION

Enthusiasm for quality improvement (QI) in healthcare continues to gather momentum, leading to increasing interest in studies of its effectiveness and to further the development of QI.[1–4] Quality improvement—often shorted to QI—involves making changes to systems and can be used in healthcare to improve specific clinical procedures (such as antibiotic prescribing) or whole care pathways.[3] QI projects are different to clinical research:[1 5–7]

they are typically pragmatic and led by local clinicians,[8] often in a single setting,[7] and they tend to be ongoing, or continual, in contrast to research which takes place over a defined period of time.[9] QI can also be conducted as research. QI research has been likened to health services research, which typically aims to examine how care can be delivered to achieve the highest quality of care,[9] and applied research, which can 'bridge the gap' (p.1962)[10] between the ideal setting of clinical trials and actual routine care. This paper explores some of the difficulties of reviewing the QI literature.

Policy makers and healthcare staff may use literature reviews to choose QI interventions wisely, thus avoiding implementing changes which are unlikely to provide any benefit.[11 12] Systematic review is one type of literature review which allows the findings and methodological quality of a large number of published articles in a scientific field to be summarised and this remains the gold standard for many decisions made in healthcare.[13] Other types of review including scoping review, narrative review, critical review and systematised review are also important to contextualise and situate research helping to provide a focal point, and to understand the current state of knowledge in a field to prevent research duplication.[14 15]

Despite the obvious importance of conducting rigorous reviews, their execution is challenging.[16] The choice of the review approach is contingent on the research question being pursued and its epistemological basis, the time available to conduct the review, the state and quality of the existing literature, and the quality standard required for the review output.[15] Once an appropriate method has been selected, rigorous literature reviewing requires adherence to a series of predefined, reproducible steps to identify, select and critically appraise relevant

research.[14] These steps are described in reporting guidelines published by the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) network (such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement),[17] but following such guidance may not always be straightforward when reviewing literature on complex interventions such as QI.[16]

Researchers deliberate how to define QI and QI interventions,[18] and QI terminology is vague, inconsistent and variable across disciplines.[19 20] Perhaps as a consequence of this, or a precursor to it, the underlying quality of QI articles and their reporting is known to be poor.[21 22] These definitional issues are not trivial problems. If a systematic review approach cannot capture *all* relevant literature because it is so difficult to classify articles as 'QI', or interventions as 'QI interventions',[19 23 24] the review will be incomplete, and less reliable estimates of the effectiveness of QI interventions may be produced.[13] Inconsistent terminology can also make it increasingly difficult for reviewers to assess studies for similarity during meta-analyses. If useful comparisons cannot be made between QI studies, knowledge synthesis becomes even more challenging.[13 25 26] Definitional problems can create similar difficulties for authors using other literature review approaches, including reduced accuracy in article identification and data extraction. Reviewers may seek precedent in the published work of others to hone their own reporting skill, but if others have been faced with inconsistencies, previous confusion can add to the cycle of poor reporting.[27]

We each used a different approach to conduct a rigorous literature review within healthcare improvement: (1) Systematic review (EJ).[21 28] (2) Integrative review (JF).[29] EJ examined QI interventions in surgery, and JF examined how improvement capability is conceptualised and assessed. A systematic review attempts to identify experimental and quasiexperimental evidence that fits prespecified eligibility criteria to answer a specific research question. It uses systematic methods designed to minimise bias.[30] An 'integrative review can precisely represent the state of the current research literature. (It) can be used to … identify the need for future research, build a bridge between related areas of work, … identify a theoretical or conceptual framework'.[31] While our studies were being conducted we were both part of a national programme with biannual learning sets where we could discuss reflections on our ongoing work. We each recorded these in our own reflective research diaries. After we had published our reviews we shared our experiences of conducting them during unstructured discussions over the telephone and by email. Despite different research fields and approaches to literature reviewing, the frustrations we encountered were surprisingly similar.

Our literature reviews required considerable debate to design a suitable search strategy, identify studies which could be defined as QI and categorise interventions into groups. In this account, we describe such debates

as a 'black box' and we attempt to unravel why debate arose and the compromises and pragmatic solutions we employed to resolve them. We have termed this a 'black box' because typically debates such as why some papers were excluded and not others, are not included in scientific reporting, but authors are encouraged to explain when much arbitration was required to resolve disputes.[32] Second, we justify the compromised and perhaps 'imperfect' decisions we made. Finally, by sharing what we learnt in the field of QI we make recommendations for others in fields characterised by similar complexity. To improve the credence of the paper, a third collaborator author (WC), challenged views of authors EJ and JF on the problems and solutions of conducting a review of literature in QI. She drew on her experiences of QI research and of writing a narrative synthesis (outside the field of QI) on the perceptions of patients and healthcare workers in women's' health.[33]

CHALLENGES OF CONDUCTING A LITERATURE REVIEW SYSTEMATICALLY IN QI

We identified two main challenges in conducting a review in QI. These were defining QI and selecting QI studies.

Defining QI

We found it challenging to delineate which literature could be described as QI. QI is an 'emerging' field of science[19 34 35] and contradicting views on how QI should be defined are commonplace in scientific abstracts[18 19 36] where a large number of terms[37] which change over time[19 20] are being used. Conceptualisation of the term 'quality improvement' is also rather unique to the healthcare field. In the engineering and manufacturing industry terms such as total quality management (TQM) and 'continuous improvement' are much more common, although how they are used to achieve the aims of QI work in healthcare may be understood differently. Thus, it is not surprising that a single accepted definition of QI is still lacking within healthcare.[38] To add to this, the volume of publications in QI in healthcare is growing,[39] and the work is scattered across different types of journals (QI journals, clinical journals or management journals)[21] and databases.[24] This may be compounded by the lack of agreement as to 'what is quality?'.[40 41] Therefore, designing a robust review strategy on 'fuzzy topics'—which are not self-defining—such as quality and QI requires the reviewer to apply judgement to show clear discernment. This may not be a problem which is unique to QI, and we therefore hope that our experiences could be useful to a wider group of researchers.

In addition, QI often uses structured techniques to support the implementation of the intervention, such as Plan-Do-Study-Act (PDSA) cycles. However, while some techniques are *related* to QI, they may not be easy to define as belonging to the field of 'QI'. For example, DMAIC (Define, Measure, Analyse, Improve, Control) is commonly used within six sigma and lean six sigma in healthcare,[42]

and we define this as belonging to the field of QI. The discipline of Instructional Systems Design also involves five integrated steps: analysis, development, design, implementation and evaluation (ADDIE)[43] and it has been used to solve specific problems in healthcare such as preventing heat loss in patients recovering from surgery.[44] However, ADDIE originates in the field of behavioural psychology and human performance technology (HPT) and is not an easy 'fit' in the field of QI. 'Quality improvement' and HPT both evolved in the context of a growing emphasis on the importance of a systematic approach to examining the quality of care.[45] However, differences in their lineage can be identified, with HPT being focused on the process of organisational learning and competency,[46] and QI being focused on studying systems and variation within systems.[47 48] Recognising the distinction and similarities between QI and other similar fields is a mindful, defensible way of discerning what 'counts' as QI.

During our reflective discussions we realised (EJ, JF) that we had both conducted a scoping review to conceptualise the term 'quality improvement'. Conceptualising QI early in the review process provided some reassurance unwanted 'noise' could be excluded and that all reviewers would consistently identify relevant literature. A third 'arbitrator' (WC) helped us to identify what questions we asked during our scoping reviews to resolve the problem of ambiguity in QI:

1. *Who should form our authorship teams?* Understanding who publishes research in QI, for example—sociologists, engineers, healthcare professionals—helped us ensure a balanced collection of viewpoints could be brought to our authorship teams. We both worked with clinicians and non-clinicians from different professional backgrounds. EJ (a physiotherapist by background) worked with medical sociologists and an anaesthetist, and JF (an engineer and healthcare manager by background) worked with experts in management and policy.

2. *What research exists on resolving ambiguities in QI?* For example, Rubenstein et al[23] classified QI articles into four types (1) Empirical literature on testing of QI. (2) QI theories and frameworks. (3) QI literature synthesis and meta-analysis. (4) Development and testing of QI-related tools. Each category has strong face validity and is easily recognised.

3. *What search terms have previous QI in healthcare authors used in their published literature reviews?* For example—the Health Foundation's QI evidence scan,[48] and Walshe[20] both review improvement methodologies using a variety of different terms. We both used a variety of different terms for 'QI' within searches to capture QI literature.[21 29] We used free-text key words to account for the problem that phrases such as 'quality improvement' and 'continuous improvement' are sometimes used synonymously.

4. *How is QI defined in the literature?*
   EJ used consensus meetings with her authorship team to reduce a list of QI definitions (box 1) to one

Box 1 Selected descriptions of QI in the literature

- ► 'QI is … a process of change in human behaviour, … driven largely by experiential learning'[87]
- ► 'Modern QI concepts had their origins in the statistical process control measurements …the input is a work process needing improvement and the output is a new, improved version of the (existing) work process'[88]
- ► 'Improvement science focuses on systematically and rigorously exploring 'what works' to improve quality in healthcare … the focus is … on structured quality improvement approaches, such as plan, do, study, act (PDSA) cycles.'[48]
- ► 'Key elements … are the combination of a 'change' (improvement) combined with a 'method' (an approach or specific tools) to attain a superior outcome.'[70]
- ► 'The combined and unceasing efforts of everyone—healthcare professionals, patients and their families, researchers, payers, planners and educators—to make the changes that will lead to better patient outcomes (health), better system performance (care) and better professional development.'[49]
- ► 'Better patient experience and outcomes achieved through changing provider behaviour and organisation through using a systematic change method'[89]

definition deemed to be clear and easy to apply to the literature. Jones et al[21] paraphrased Batalden and Davidoff[49] to define QI as: 'deliberate structured process of purposeful efforts to make changes that will lead to better outcomes, better system performance, and better professional development'.

In addition to defining QI, JF needed to define 'improvement capability' to isolate papers for her review topic. Similarly, literature review revealed no accepted definition. Therefore, Furnival et al (2017)[29] used Whittemore and Knafl's integrative review method[50] which allows for several perspectives of a topic to be used. Improvement capability was defined as: 'the organisational ability to intentionally and systematically use improvement approaches, methods and practices, to change processes and products/services to generate improved performance.'[29]

5. *How is QI distinct from, and how does it overlap with, other fields?* To define the field of QI we worked with experts in QI and related fields. EJ adopted values and norms unique to QI in healthcare, distinguishing it as a separate field in order to deploy a specific review methodology in the field of surgery. JF positioned QI *across* fields, by examining how QI is used in healthcare based on its heritage from manufacturing and industry, partly because JF had a background in engineering, and because there is so much to learn from other fields. The resultant search sought to identify qualitative and quantitative instruments that had been tested across fields and all sectors were included. No date restrictions were used.
   EJ contacted experts in related fields such as Human Factors to identify which features could be used to distinguish QI papers and human factors papers

from each other. These discussions resulted in a co-authored manuscript with Human Factors and QI experts to expose the similarities and differences between the fields.[51] Conversely, JF acknowledged that HFs (Human Factors) and ergonomics is an integral part of a QI specialist's skill set, and as such a definition of how QI overlaps (rather than being distinct from) other fields became more important.

6. *We conducted pilot literature searches* to ensure a selection of known surgical QI papers[42] and improvement capability papers[52–54] were successfully captured. This verified that EJ's choice of terms could capture papers which conceptualised QI as relating to a set of values and norms in the healthcare field, and that JF's choice of terms could capture papers published in both healthcare and related fields.

The search strategy and databases used by EJ and JF are included in online supplementary file appendix 1.

The methods we used to conceptualise the term 'quality improvement' have some limitations. For example, asking experts to clarify the nature of QI is subject to heterogeneity in opinions. We also made some concessions. For example, EJ used broad search terms such as 'quality adj2 improve$' (where the terms 'quality' and variations of 'improve' such as 'improvement' are identified within two words of each other). This captured titles such as 'Can quality circles improve hospital-acquired infection control?',[55] which would not otherwise be identified by 'quality improvement'. However, this term also captured articles which described quality of care, rather than QI. To overcome problems with specificity, the Medical Subject Headings (MeSH) 'Quality Improvement/' and 'TQM' are now in use, which guard against the possibility that review teams may not have thought of all relevant synonyms.[56]

Additionally, our strategies to resolve ambiguity did not produce perfect classifications. In a relatively new field challenged by inconsistent conceptual and terminological definition, QI can mean different things to different people, depending on professional background, and research aims and intentions. However, documenting which compromises were made and presenting these decisions to others in a discerning, mindful way provides further opportunity for learning which otherwise might be missed. For example, EJ published the review protocol,[28] and registered it with PROSPERO[57] which allows readers to compare the published review with the protocol to check all data were extracted as intended. We both kept a reflective account of decision making. JF used her reflective account to explain that variable conceptualisation and assessment of capability could be eased by employing the integrative method which allows for more than one a priori definition to be used during searching.

Selecting suitable studies in the QI field

Accurately selecting QI studies during a literature review is problematic because it is hard to determine what makes an intervention a 'QI intervention'.[18] At first glance, any clinical healthcare intervention including pharmaceutical, surgical, physiotherapeutic interventions, for example, could be classed as 'quality improvement'. All of these interventions aim to improve outcome (quality of life or length of stay, eg), which could in turn impact quality of care. Likewise, checklists, algorithms or pathways are systematic clinical activities that aim to improve outcome. For example, Enhanced Recovery After Surgery (ERAS) pathways combine early mobilisation, intraoperative fluid balance and carbohydrate loading, to achieve an overall effect (reducing surgical stress response to reduce length of stay) which improves a clinical process (surgery).[58–62] Like ERAS pathways, QI work tends to involve the implementation of multiple interventions to improve clinical processes, but ERAS is not routinely referred to as a QI intervention in the literature.

Poor conceptualisation of what makes an intervention a QI intervention prompted EJ to follow the recommendations of Shepperd *et al*,[16] who suggest that known taxonomies (or rule sets) could aid intervention classification. Jones *et al*[21] used a taxonomy produced by Shojania *et al*[63] (table 1). This taxonomy was built on other well-established taxonomies of behaviour change interventions[64–68] after examining QI across a range of medical specialities.[69] Shojania *et al*[63] describe a QI intervention as something which can improve processes to *support* clinical activity (such as a reminder system to wash your hands). Therefore, not all interventions to improve quality of care are quality interventions and the content of care (such as a new surgical technique) would not be classed as a QI intervention.[9]

EJ modified Shojania's original taxonomy to stipulate that a QI intervention must be supported by a QI technique (such as a PDSA cycle) because QI techniques are often cited in the QI literature.[48 70] Not all QI interventions require the use of a QI technique, but EJ and her authorship team used the wider literature to create a pragmatic operationalisation of what constitutes QI.[21]

Conversely, JF did not search for interventions, but frameworks and instruments for assessing or measuring improvement capability. However, JF faced similar challenges due to wide heterogeneity in the framework and instrument constructs. Rather than starting out with a narrow conceptualisation of QI and improvement capability, a broader literature was targeted. Keywords were selected to take account of the broad array of terminological heterogeneity across many disciplines, rather than a taxonomy of categories. Therefore this searching strategy required supplementation to identify further high-quality references.[71 72] JF searched across several bibliographical databases with thesaurus terms, MeSH terms and broad-based terms. JF also used citation checking and reference searching to 'snowball' iteratively searching backwards and forwards on citations to find more obscure yet relevant articles until saturation was reached.[73] JF operationalised inclusion criteria that improvement capability assessment instruments could only be included within the sample if they had been evaluated at least once and were

Table 1 Quality improvement taxonomy

	QI strategy	Definition	Example methods	Surgical examples
Articles reporting any QI intervention (1–9) must include one additional item (10–11)				
1	Provider reminder systems	Any 'clinical encounter-specific' information intended to prompt a clinician to recall information or consider a specific process of care	Decision aids Reminders	MEWS The WHO Surgical Safety Checklist
2	Facilitated relay of clinical data to providers	Transfer of clinical information from patients to the provider (not during a patient visit)	Telephone call Postal contact	Relay of BP measurements to the preassessment team
3	Provider education	Dissemination of information	Educational outreach visits Distribution of educational material Clinical guideline information	Component separation, training and recurrence rates Cadaveric training and surgeon confidence
4	Patient education	Dissemination of information	Distribution of educational material Individual or group sessions	Trimodal pre-habilitation programmes compliance and effect on LOS
5	Promotion of self-management	Access to a resource that enhances the patient's ability to manage their condition	BP devices Patient diaries	Follow-up phone calls with recommended adjustments to care
6	Patient reminders	Any methods of encouraging patient compliance to self-management	Appointment reminders	SMS exercise reminders before bariatric surgery
7	Organisational change	Any change in organisational structure	Multidisciplinary teams Communication Health records	Changes to staff rota to facilitate early patient mobilisation after elective arthroplasty
8	Financial, regulatory, or legislative incentives	Any financial bonus, reimbursement or provider licensure scheme	Positive or negative incentives for providers or patients.	18-week wait target for elective orthopaedic surgery
9	Feedback	Any feedback of clinical performance	Distribution of feedback via staff education sessions or emails. Can occur as part of SPC or audit and feedback	Percentage of patients achieving target LOS
QI techniques (10–11)				
10	Audit and feedback	Any feedback of clinical performance summarising percentages of patients who have achieved a target outcome which has been measured at intervals over time	PROMs LOS Morbidity and mortality	Percentage of patients achieving target LOS
11	QI methods	Systematic techniques for identifying defects in clinical systems and making improvements, typically involving process measurement and remeasurement	PDSA, Six Sigma, TQM, CQI, SPC, Lean	Improving processes for acetabular cup placement in minimally invasive hip surgery

Continued

Table 1	Continued		
QI strategy	**Definition**	**Example methods**	**Surgical examples**

Adapted from Shojania *et al* (2004) *Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies* (Vol. 1: Volume 1—Series Overview and Methodology). Technical Reviews, Rockville (Maryland): Agency for Healthcare Research and Quality (USA).

BP, Blood Pressure; CQI, Continuous Quality Improvement; LOS, Length of Stay; MEWS, Modified Early Warning System; PDSA, Plan-Do-Study-Act; PROM, Patient Reported Outcome Measure; QI, quality improvement; SMS, Short Message Service; SPC, Statistical Process Control; TQM, total quality management.

supported by empirical data. Multiple uses or iterations of the same instrument were discarded.

DISCUSSION

Our report provides a reflexive, critical overview of the process of reviewing the literature in the field of QI. Through clear documentation of reproducible methods of selecting articles for inclusion (such as using a known taxonomy or method) threats to the validity of our results could be reduced. However, the imperfections of these approaches are well known. Even when a research aid such as taxonomy has been selected, they can be challenging to apply, especially when the terminology used is inconsistent and heterogeneous.[36 37 74] For example, Colquhoun *et al*[75] explain that one author may use the behaviour change wheel[76] to refer to an intervention as 'environmental restructuring', and another may use the EPOC (Cochrane Effective Practice and Organisation of Care) framework[77] to describe the same thing as a 'structural intervention'. This can exaggerate confusion on how to select and implement the most appropriate taxonomy from the many aids to classification that exist in the literature.[12] Also, the terminology used in Shojania *et al*'s[63] taxonomy is not always consistent with the wider QI literature. While more time could have been invested to develop a new set of criteria, the divided opinions of researchers in the field may still have threatened its implementation. Therefore, developing a new taxonomy would have been an 'over-engineered' solution which would not necessarily be more helpful than what Jones *et al*[21] were able to produce. Asking the experts whether they judge their work to sit within the field of QI is also an imperfect indicator of how QI should be defined, yet, the literature is also imperfect.

The methods we suggest to identify scientific publications about QI are not infallible, but they are defensible. There is a large amount of heterogeneity between QI articles in terminology and methodology (diversity in intervention and study design).[42 78–81] A 'virtually infinite number of combinations of features and local environmental circumstances' (p.244)[82] also makes the synthesis of QI evidence extremely challenging. Additionally, terms such as QI and improvement capability do not have unproblematic widely agreed meanings. This necessitated judgements to be discerned from a variety of sources including existing literature, expert opinion and personal experience, but these judgements may not be universally accepted within the QI community. For example, EJ's judgement that QI interventions do not include clinical interventions may be contentious for some, but based on the evidence available and epistemological standpoint, this was deemed wholly appropriate to satisfy the aims of the systematic review.

Recognising the value of a more plural epistemological outlook may also be important for authors reviewing the literature in complex or relatively new scientific fields. EJ followed a systematic review approach which is commonly associated with a positivist approach to gathering and synthesising knowledge. This was well suited to her research question: 'what is the completeness of reporting of QI interventions in the perioperative literature?'. Conversely, JF adopted a constructivist perspective asking, 'how is improvement capability conceptualised and assessed?'. While different review questions can drive the method selected for reviewing, the question alone cannot generate a typology of review method.[83] We found value in exploring post hoc the different viewpoints from which we designed our review protocols, recognising their strengths, rather than polarities. Indeed, scholars in the field of QI itself argue that hard scientific evidence needs to be combined with an understanding of beliefs and 'soft intelligence' to ensure successful interventions can be spread and adopted within complex organisations such as the National Health Service.[84 85]

The problems we describe cannot be resolved immediately but methodological direction is developing.[82] For example, Colquhoun *et al*,[37] through an international working group, are investigating whether a single consensus on the taxonomic classification of interventions (including QI interventions) is possible. While in clinical practice and research, a degree of uncertainty and creativity can be celebrated to allow thinking to evolve,[86] we suggest that it is important for the QI community as a whole to seek agreement on what constitutes the term QI for two key reasons. First, we must ensure that patients can gain the most benefit from systematic reviews of QI work. To translate the findings of QI into practice, reviewers must be able to identify the QI literature and accurately synthesise it so that robust conclusions may be drawn. Second, agreement on what constitutes QI is needed so that QI research capability can be better facilitated. Organisations offering funding or guidance, may not be able to standardise eligibility for support if they struggle to identify which activities count as improvement and which do not. This may in time relieve the financial and social burden of poor-quality work.

Recommendations

After our reviews were published, we discussed the challenges we faced and how we resolved them during a series of reflective discussions by telephone and email. Through arbitration with a third author (WC), we were able to suggest what other authors could do to ease the process of literature-reviewing in complex fields. Our approach does not offer a definitive solution, but rather a starting point which should be critiqued and refined depending on the review question and aims. Our approach is based on our own reflections of using a systematic approach to reviewing literature in a complex field. There are many existing frameworks and resources (such as the PRISMA statement) to aid review authors, but by using the specific example of QI in healthcare we add to this literature, providing practical examples from our own experience. During our own work, in a field where there is much contention not just about what QI is, but where it fits in a landscape of other sciences, we made pragmatic choices to complete our reviews with finite time and resources. Therefore, we hope that our recommendations can be used as a starting point for authors in similar complex fields to strengthen the practice of systematic and literature reviews:

1. Maintain transparency and be as thorough as possible (using supplementary materials if required) to expose the challenges you faced to allow the reader to make judgements about the defensibility of your methodological choices.
2. Assemble a multiprofessional review team (our teams included clinicians, social scientists and engineers).
3. Ensure the review team can communicate in an open and honest way about uncertainties and compromises.
4. Conduct a literature scoping exercise to resolve ambiguities in defining the topic and key terms before starting the main review.
5. Apply a taxonomy or rule to aid classification of concepts.
6. Contact experts in related fields and authors of papers selected for full-text review to clarify whether a paper meets inclusion criteria if the literature review team cannot agree and ambiguity exists.
7. Maintain documentation of reflexivity and make this available to experts in your field.
8. Consider plurality as an approach when working with a multidisciplinary review team.

CONCLUSION

The QI community should work together as a whole to create a scientific field with a shared vision of what QI is and how literature can be accurately identified to sit within the field of QI . This in turn will allow successful application of solutions, such as those presented in this paper, to enable faster and more accurate identification and synthesis of QI evidence. This will ensure that effective QI work can be adopted rapidly and reliably with greatest impact for patient care. Our recommendations could also be helpful for systematic reviewers wishing to evaluate complex interventions in similar fields.

REFERENCES

1. Dixon-Woods M, Martin GP. Does quality improvement improve quality? *Future Hosp J* 2016;3:191–4.
2. Koetsier A, van der Veer SN, Jager KJ, *et al*. Control charts in healthcare quality improvement. A systematic review on adherence to methodological criteria. *Methods Inf Med* 2012;51:189–98.
3. Powell AE, Rushmer RK, Davies HT. A systematic narrative review of quality improvement models in health care. In: *NHS quality improvement Scotland*, 2009.
4. Steinman MA, Ranji SR, Shojania KG, *et al*. Improving antibiotic selection: a systematic review and quantitative analysis of quality improvement strategies. *Med Care* 2006;44:617–28.
5. Portela MC, Pronovost PJ, Woodcock T, *et al*. How to study improvement interventions: a brief overview of possible study types. *BMJ Qual Saf* 2015;24:325–36.
6. Sobo EJ, Bowman C, Gifford AL. Behind the scenes in health care improvement: the complex structures and emergent strategies of implementation science. *Soc Sci Med* 2008;67:1530–40.
7. Øvretveit J. Producing useful research about quality improvement. *Int J Health Care Qual Assur* 2002;15:294–302.
8. M J, Thorlby R. High quality care for all: Briefing on NHS next stage review final report. In: *The Kings Fund*, 2008: 1–8.
9. Margolis P, Provost LP, Schoettker PJ, *et al*. Quality improvement, clinical research, and quality improvement research--opportunities for integration. *Pediatr Clin North Am* 2009;56:831–41.
10. Ting HH, Shojania KG, Montori VM, *et al*. Quality improvement: science and action. *Circulation* 2009;119:1962–74.
11. England NHS. *Transforming urgent and emergency care services in England: urgent and emergency care review end of phase 1 report*. United Kingdom: Leeds, 2013.
12. Dijkers MP. Reporting on interventions: issues and guidelines for rehabilitation researchers. *Arch Phys Med Rehabil* 2015;96:1170–80.
13. York Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*, 2009: 1–281.
14. Grant MJ, Booth A. A typology of reviews : an analysis of 14 review types and. *Health Info Libr J* 2009;26:91–108.
15. Ridley D. *The literature review: a step-by-step guide for students*. 2nd ed. London: SAGE, 2012.
16. Shepperd S, Lewin S, Straus S, *et al*. Can we systematically review studies that evaluate complex interventions? *PLoS Med* 2009;6:e1000086.
17. Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010;8:336–41.
18. Shojania KG, Ranji SR, McDonald KM, *et al*. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA* 2006;296:427–40.
19. Hempel S, Rubenstein LV, Shanman RM, *et al*. Identifying quality improvement intervention publications - A comparison of electronic search strategies. *Implement Sci* 2011;6:1–10.

20. Walshe K. Pseudoinnovation: the development and spread of healthcare quality improvement methodologies. *Int J Qual Health Care* 2009;21:153–9.

21. Jones EL, Lees N, Martin G, *et al*. How well is quality improvement described in the perioperative care literature? A systematic review. *Jt Comm J Qual Patient Saf* 2016;42:196–AP10.

22. Leviton L. Reconciling complexity and classification in quality improvement research. *BMJ Qual Saf* 2011;20(Suppl 1):i28–9.

23. Rubenstein LV, Hempel S, Farmer MM, *et al*. Finding order in heterogeneity: types of quality-improvement intervention publications. *Qual Saf Health Care* 2008;17:403–8.

24. Hoffmann T, Erueti C, Thorning S, *et al*. The scatter of research: cross sectional comparison of randomised trials and systematic reviews across specialties. *BMJ* 2012;344:e3223.

25. Shrier I, Boivin J-F, Platt RW, *et al*. The interpretation of systematic reviews with meta-analyses: an objective or subjective process? *BMC Med Inform Decis Mak* 2008;8:19.

26. Popay J, Roberts H, Sowden A, *et al. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews - A Product from the ESRC Methods Programme*, 2006.

27. Conn VS. Unpacking the black box: countering the problem of inadequate intervention descriptions in research reports. *West J Nurs Res* 2012;34:427–33.

28. Jones E, Lees N, Martin G, *et al*. Describing methods and interventions: a protocol for the systematic analysis of the perioperative quality improvement literature. *Syst Rev* 2014;3.

29. Furnival J, Boaden R, Walshe K. Conceptualizing and assessing improvement capability: a review. *Int J Qual Heal Care* 2017;29:604–11.

30. Higgins J, Green S. *Cochrane Handbook for systematic reviews of interventions version 5.1.0*, 2011.

31. Russell CL. An overview of the integrative research review. *Prog Transplant* 2005;15:8–13.

32. Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339.

33. Carter W, Bick D, Mackintosh N, *et al*. A narrative synthesis of factors that affect women speaking up about early warning signs and symptoms of pre-eclampsia and responses of healthcare staff. *BMC Pregnancy Childbirth* 2017;17:1–16.

34. Ogrinc G, Davies L, Goodman D, *et al*. Squire 2.0 (*standards for quality improvement reporting excellence)*: revised publication guidelines from a detailed consensus process. *BMJ Qual Saf* 2016;25:986–92.

35. Dixon-Woods M, Bosk CL, Aveling EL, *et al*. Explaining Michigan: developing an ex post theory of a quality improvement program. *Milbank Q* 2011;89:167–205.

36. McDonald K, Schultz E, Chang C. *Methods research report: through the quality kaleidoscope: reflections on the science and practice of improving health care quality methods research report through the quality kaleidoscope: reflections on the science and practice of improving health care*. Agency Healthc Res Qual, 2013.

37. Colquhoun H, Leeman J, Michie S, *et al*. Towards a common terminology: a simplified framework of interventions to promote and integrate evidence into health practices, systems, and policies. *Implement Sci* 2014;9:51.

38. Kaplan HC, Brady PW, Dritz MC, *et al*. The influence of context on quality improvement success in health care: a systematic review of the literature. *Milbank Q* 2010;88:500–59.

39. Stelfox HT, Palmisani S, Scurlock C, *et al*. The "To Err is Human" report and the patient safety literature. *Qual Saf Health Care* 2006;15:174–8.

40. Swinglehurst D, Emmerich N, Maybin J, *et al*. Confronting the quality paradox: towards new characterisations of 'quality' in contemporary healthcare. *BMC Health Serv Res* 2015;15:240.

41. Boaden R, Furnival J. Quality improvement in healthcare. In: Walshe K, Smith J, eds. *Health care management*. 3rd ed. England: Open University Press, 2016: p. 454–478.

42. Nicolay CR, Purkayastha S, Greenhalgh A, *et al*. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. *Br J Surg* 2012;99:324–35.

43. O'Driscoll T. Chronicling the emergence of human performance technology. *Perf Improv* 2015;54:34–47.

44. Hayes JS, Tyler-Ball S, Cohen SS, *et al*. Evidence-Based practice and heat loss prevention in trauma patients. *J Nurs Care Qual* 2002;16:13–16.

45. Marjoua Y, Bozic KJ. Brief history of quality movement in US healthcare. *Curr Rev Musculoskelet Med* 2012;5:265–73.

46. Jang HY. Reconsidering human performance technology. *Perf Improv* 2008;47:25–33.

47. Grol R, Baker R, Moss F. Quality improvement research: understanding the science of change in health care. *Qual Saf Health Care* 2002;11:110–1.

48. The Health Foundation. Report: improvement science. London Research Scan; 2011.

49. Batalden PB, Davidoff F. What is "quality improvement" and how can it transform healthcare? *Qual Saf Health Care* 2007;16:2–3.

50. Whittemore R, Knafl K. The integrative review: updated methodology. *J Adv Nurs* 2005;52:546–53.

51. Hignett S, Jones EL, Miller D, *et al*. Human factors and ergonomics and quality improvement science: integrating approaches for safety in healthcare. *BMJ Qual Saf* 2015;24:250–4.

52. Bessant J, Francis D. Developing strategic continuous improvement capability. *Int Jrnl of Op Prod Mnagemnt* 1999;19:1106–19.

53. Anand G, Ward PT, Tatikonda MV, *et al*. Dynamic capabilities through continuous improvement infrastructure. *J Oper Manage* 2009;27:444–61.

54. Adler PS, Riley P, Kwon S-W, *et al*. Performance improvement capability: keys to accelerating performance improvement in hospitals. *Calif Manage Rev* 2003;45:12–33.

55. Forster DH, Krause G, Gastmeier P, *et al*. Can quality circles improve hospital-acquired infection control? *J Hosp Infect* 2000;45:302–10.

56. National Libraray of Medicine. Pubmed expert searching: using PubMed to get advanced results, 2006. Available: https://nnlm.gov/training/resources/pubmedalt.html [Accessed 11 Jan 2014].

57. Jones E, Lees N, Martin G, *et al*. Describing methods and interventions: a protocol for the systematic analysis of the perioperative quality improvement literature. *Syst Rev* 2014;3:98.

58. Khan S, Wilson T, Ahmed J, *et al*. Quality of life and patient satisfaction with enhanced recovery protocols. *Color Dis* 2010;12:1175–82.

59. Jones EL, Wainwright TW, Foster JD, *et al*. A systematic review of patient reported outcomes and patient experience in enhanced recovery after orthopaedic surgery. *Ann R Coll Surg Engl* 2014;96:89–94.

60. Hall TC, Dennison AR, Bilku DK, *et al*. Enhanced recovery programmes in hepatobiliary and pancreatic surgery: a systematic review. *Ann R Coll Surg Engl* 2012;94:318–26.

61. Lv D, Wang X, Shi G. Perioperative enhanced recovery programmes for gynaecological cancer patients (review). *Cochrane database Syst Rev* 2010;(6).

62. Spanjersberg WR, Reurings J, Keus F, *et al*. Fast track surgery versus conventional recovery strategies for colorectal surgery. *Cochrane Database Syst Rev* 2011;90.

63. Shojania K, McDonald K, RM W, *et al. Closing the quality gap: a critical analysis of quality improvement strategies. (Vol. 2: diabetes Mellitius care) technical review 9 (contract No. 290-02-0017 to the Stanford University–UCSF evidence-based practice center*. AHRQ Publ No 04-0051-2 Rockville, MD Agency Healthc Res Qual US Dep Heal Hum Serv, 2004.

64. Ellis P, Robinson P, Ciliska D. *Difiusion and dissemination of evidence based cancer control interventions: summary. Evidence Report/Technology assessment No. 79*. Rockville, MD, US: Agency for Healthcare Research and Quality, 2003.

65. Grimshaw J, Shirran L, Thomas R, *et al*. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care* 2001;39(Suppl 2):112–45.

66. Stone EG, Morton SC, Hulscher ME, *et al*. Article Interventions That Increase Use of Adult Interventions that increase use of adult Immunization and Cancer Screening Services : A Meta-Analysis. *Ann Intern Med* 2002;136:641–51.

67. Stafford RS. Alternative strategies for controlling rising cesarean section rates. *JAMA* 1990;263:683–7.

68. Weingarten SR, Henning JM, Badamgarav E, *et al*. Interventions used in disease management programmes for patients with chronic illness-which ones work? meta-analysis of published reports. *BMJ* 2002;325:925.

69. McDonald KM, Schultz EM, Chang C. Evaluating the state of quality-improvement science through evidence synthesis: insights from the closing the quality gap series. *Perm J* 2013;17:52–61.

70. The Health Foundation. *Quality improvement made simple. what every board should know about healthcare quality improvement*. London, 2010.

71. Shaw RL, Booth A, Sutton AJ, *et al*. Finding qualitative research: an evaluation of search strategies. *BMC Med Res Methodol* 2004;4:5.

72. Papaioannou D, Sutton A, Carroll C, *et al*. Literature searching for social science systematic reviews: consideration of a range of search techniques. *Heal Info Libr J* 2009;27:114–22.

73. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 2005;331:1064–5.

74. Lokker C, McKibbon KA, Colquhoun H, *et al*. A scoping review of classification schemes of interventions to promote and integrate evidence into practice in healthcare. *Implement Sci* 2015;10:1–12.

75. Colquhoun H, Leeman J, Michie S, *et al*. Additional File 1. The mapping of : Behaviour Change Wheel, EPOC categories, Leeman taxonomy, and behaviour change techniques. In: Towards a common terminology: a simplified framework of interventions to promote and integrate evidence into health practic. *Implement Sci* 2014;9.

76. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Sci* 2011;6.

77. Shepperd S, Lewin S. *On behalf of the effective practice and organisation of care group. The effective practice and organisation of care (EPOC) taxonomy of health systems interventions. EPOC resources for review authors*. Oslo: Norwegian: Knowledge Centre for the Health Services, 2015: 1–11.

78. Thor J, Lundberg J, Ask J, *et al*. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care* 2007;16:387–99.

79. Treadwell JR, Lucas S, Tsou AY. Surgical checklists: a systematic review of impacts and implementation. *BMJ Qual Saf* 2014;23:299–318.

80. Howell A-M, Panesar SS, Burns EM, *et al*. Reducing the burden of surgical harm: a systematic review of the interventions used to reduce adverse events in surgery. *Ann Surg* 2014;259:630–41.

81. Portela MC, Pronovost PJ, Woodcock T, *et al*. How to study improvement interventions: a brief overview of possible study types. *Postgrad Med J* 2015;91:343–54.

82. Alexander JA, Hearld LR. What can we learn from quality improvement research? A critical review of research methods. *Med Care Res Rev* 2009;66:235–71.

83. Gough D, Thomas J, Oliver S. Clarifying differences between review designs and methods. *Syst Rev* 2012;1:1–9.

84. Perla RJ, Parry GJ. The epistemology of quality improvement: it's all Greek. *BMJ Qual Saf* 2011;20(Suppl 1):i24–7.

85. Martin GP, Aveling E-L, Campbell A, *et al*. Making soft intelligence hard: a multi-site qualitative study of challenges relating to voice about safety concerns. *BMJ Qual Saf* 2018;27:710–7.

86. Engebretsen E, Heggen K, Wieringa S, *et al*. Uncertainty and objectivity in clinical decision making: a clinical case in emergency medicine. *Med Health Care Philos* 2016;19:595–603.

87. Ogrinc G, Mooney SE, Estrada C, *et al*. The Squire (standards for quality improvement reporting excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care* 2008;17(Suppl 1):i13–32.

88. Solberg LI, Mosser G, McDonald S. The three faces of performance measurement: improvement, accountability, and research. *Jt Comm J Qual Improv* 1997;23:135–47.

89. Øvretveit J. *Does improving quality save money? A review of evidence of which improvements to quality reduce costs to healthcare providers*. London: Health Foundation, 2009.

9