EDITOR'S CHOICE

Identifying continuous quality improvement publications: what makes an improvement intervention 'CQI'?

Sean M O'Neill,[1,2] Susanne Hempel,[1] Yee-Wei Lim,[1] Marjorie S Danz,[1,3] Robbie Foy,[4] Marika J Suttorp,[1] Paul G Shekelle,[1,3,5] Lisa V Rubenstein[1,3,5]

[1]RAND Corporation, Santa Monica, California, USA
[2]Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA
[3]Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, California, USA
[4]Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
[5]David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA

Correspondence to
Sean M O'Neill, Northwestern University, Feinberg School of Medicine, Chicago, IL 60611, USA; soneill@fsm.northwestern.edu

Accepted 23 May 2011
Published Online First 4 July 2011

ABSTRACT

Background: The term continuous quality improvement (CQI) is often used to refer to a method for improving care, but no consensus statement exists on the definition of CQI. Evidence reviews are critical for advancing science, and depend on reliable definitions for article selection.

Methods: As a preliminary step towards improving CQI evidence reviews, this study aimed to use expert panel methods to identify key CQI definitional features and develop and test a screening instrument for reliably identifying articles with the key features. We used a previously published method to identify 106 articles meeting the general definition of a quality improvement intervention (QII) from 9427 electronically identified articles from PubMed. Two raters then applied a six-item CQI screen to the 106 articles.

Results: Per cent agreement ranged from 55.7% to 75.5% for the six items, and reviewer-adjusted intra-class correlation ranged from 0.43 to 0.62. 'Feedback of systematically collected data' was the most common feature (64%), followed by being at least 'somewhat' adapted to local conditions (61%), feedback at meetings involving participant leaders (46%), using an iterative development process (40%), being at least 'somewhat' data driven (34%), and using a recognised change method (28%). All six features were present in 14.2% of QII articles.

Conclusions: We conclude that CQI features can be extracted from QII articles with reasonable reliability, but only a small proportion of QII articles include all features. Further consensus development is needed to support meaningful use of the term CQI for scientific communication.

INTRODUCTION

Continuous quality improvement (CQI) represents a set of methods for improving healthcare[1–4] that originated from industrial process improvement approaches.[5 6] One evidence review describes CQI as 'a philosophy of continual improvement of the processes associated with providing a good or service that meets or exceeds customer expectations'.[7] Although a useful starting point, this definition has not emerged from formal consensus processes, has not been tested for reliability, and may therefore be difficult to operationalise in evidence syntheses. Greater consensus on key features of CQI that could be reliably operationalised would improve the reporting, cataloguing, and systematic review of CQI interventions.

We acknowledge that meanings fluctuate over time.[8 9] The term CQI has a complex heritage from use in both industry and healthcare, and seeking to create a normative definition may perturb this evolution.[9] Science, however, depends upon clear word usage for communication, and efforts to understand scientific meaning have often promoted scientific development in both clinical[10–12] and methodological[13–19] domains. In the work presented here, we aimed to understand the current usage of the term CQI as a step towards improving scientific communication in the quality improvement field.

This work is part of the 'Advancing the Science of Continuous Quality Improvement' (ASCQI) Program funded by the Robert Wood Johnson Foundation, a US-based healthcare-oriented philanthropic organisation. One ASCQI aim was to 'develop methods, tools and standards for the design, conduct and reporting of CQI research and evaluations, including standardised typologies, definitions and measures of key concepts and consensus statements'.[20]

Towards that aim, this study developed a screen for CQI features, tested it for reliability, and applied it to electronically identified quality improvement intervention (QII) articles to assess which key CQI features are most commonly present in today's quality improvement literature.

METHODS

Overview

We first elicited a broad range of existing definitions for CQI, distilled them into candidate key features, and engaged an expert panel to rate and refine the features. We then used a previously published QII definition[21] as the basis for a QII screening form and applied it to articles from a broad electronic search. Finally, we operationalised the highest-scoring consensus-based CQI features as an assessment form and applied it to the QII article set.

Identification of potential key features of CQI

To identify key features of CQI, we conducted a simplified, sequential group consensus process, similar to a repeated focus group with feedback. We organised a 12-member expert panel, intentionally encompassing a diverse range of methodological perspectives representing both quality improvement and research expertise. Individual experts included process consultants, researchers and institutional decisionmakers from both the USA and the UK; several additionally serve as editors of clinical or quality improvement journals (see 'Acknowledgements' for complete list). To begin generating potentially definitional features of CQI, Robert Wood Johnson Foundation staff reviewed grant applications to the ASCQI Program and abstracted 48 phrases used by applicants to define 'CQI'. Two authors (LR, SH) independently reviewed these phrases to ascertain common themes, reconceptualised them as a list of unique, potentially definitional features, and then met to discuss and reach agreement on the list.

The expert panel then completed an online survey of the features, reviewed survey results, and discussed the results on two conference calls. The survey asked, for each feature: 'Is this feature necessary (definitional) for CQI?' (5=definitely; 4=probably; 3=no difference; 2=probably not; 1=definitely not). The survey and discussion process enabled the addition of features to the original set from other sources, as suggested by the panel or research team.[20–29] Table 1 lists 12 features (A–L) finalised when the process had ceased generating additional potential features. Panelists rated the 12 features again at a final in-person meeting, resulting in six features (A, C, D, E, G, K) rated as 'definitely' or 'probably' necessary (definitional) for CQI (median value ≥4.0). The final column in table 1 shows which

Table 1 Potentially definitional continuous quality improvement (CQI) features

Feature	Description	'Definitely' or 'probably' definitional for CQI	Item(s) on CQI features assessment form
A	The intervention involves an iterative development and testing process such as PDSA (Plan-Do-Study-Act)	X	CQI-1, CQI-5
B	The intervention is designed and/or carried out by teams		
C	The intervention uses systematic data-guided activities to achieve improvement	X	CQI-3, CQI-5
D	The intervention involves feedback of data to intervention designers and/or implementers	X	CQI-2, CQI-5
E	The intervention aims to change how care is organised, structured, or designed	X	QII-4
F	The intervention aims to change the daily work or routine within an organisation		
G	The intervention identifies one or more specific methods (eg, change strategies) aimed at producing improvement	X	CQI-4
H	The intervention aims to redesign work processes		
I	The intervention uses available previously established evidence relevant to the target QI problem or goal		
J	The intervention seeks to create a culture or mindset of quality improvement		
K	The intervention is designed/implemented with local conditions in mind	X	CQI-6
L	The intervention is shaped by clearly defined desired outcomes/targets		

items on the final CQI features assessment form reflected each 'definitely' or 'probably' definitional feature.

Criteria for identifying QII studies

We focused on QII studies that, as described previously,[21] addressed effectiveness, impacts, or success; qualitative, quantitative and mixed-methods studies were all considered eligible. Our QII screening form identified articles that: (QII-1) reported on an intervention implemented in or by a healthcare delivery organisation or organisational unit; (QII-2) reported qualitative or quantitative data on intervention effectiveness, impacts, or success; (QII-3) reported on patient (or care giver) health outcomes; and (QII-4) aimed to change how delivery of care was routinely structured within a specific organisation or organisational unit. All four QII criteria had to be present according to two independent reviewers for an article to be included. We thus excluded studies that only reported cost or provider knowledge/attitude measures.[21] The fourth criterion, QII-4, conceptually overlaps with potential CQI feature E ('The intervention aims to change how care is organised, structured, or designed') as identified by our CQI expert panel. Because we selected articles for our QII sample based on this criterion, 100% of studied articles had this feature.

Criteria for assessing CQI features

With feature E already part of the QII screening form, we then incorporated the five remaining 'definitely' or 'probably' definitional CQI features into a six-item assessment form (table 2), and refined the form and its guidelines through pilot testing. These five features were CQI-1 ('The intervention involves an iterative development and testing process such as PDSA (Plan-Do-Study-Act)'), CQI-2 ('The intervention involves feedback of data to intervention designers and/or implementers'), CQI-3 ('The intervention uses systematic data-guided activities to achieve improvement'), CQI-4 ('The intervention identifies one or more specific methods (eg, change strategies) aimed at producing improvement') and CQI-6 ('The intervention is designed/implemented with local conditions in mind'). The concept of being 'data driven' was a consistent theme of the expert panel discussions, manifested through items CQI-1, CQI-2 and CQI-3, which all reflect data use but do not use the term 'data driven'. We added CQI-5 as a potentially more direct assessment.

We used a three-point scale with explicit criteria for all items during pilot testing. However, CQI-5 (data driven) and CQI-6 (designed for local conditions) were not reliable in this form. We therefore used a five-point implicit (reviewer judgement-oriented) review scale for these two items. Pilot testing showed better reliability for the three-point and five-point scales than simple yes/no responses.

Exploratory items

To further enhance our understanding of the QII and CQI literature, we collected additional information on reviewed articles. We assessed setting, evaluation target (ie, change package, change method, or both), evaluation design (ie, presence/absence of a comparison group), researcher involvement in authorship, results (ie, whether the intervention demonstrated positive effects), and journal type (to explore potential differences in reporting across publication venues). 'Change package' describes the set of specific changes for improving care (reminders, tools, or other care model or prevention elements) implemented in a QII, while 'change method' describes the approach used to introduce and implement the change package (eg, CQI, Lean, Six-Sigma, Reengineering). For assessing journal type, we characterised journals as clinical (general, nursing, or specialty) or quality improvement/health services research.

QII sample identification and screening

To reflect usual methods for evidence review, we began with electronically searched articles. We developed search strategies for the MEDLINE (Ovid) and PubMed databases based on free text words, medical subject headings, QI intervention components, CQI methods, and combinations of the strategies (Hempel *et al*, submitted). Searches included a broad range of terms ('quality' AND 'improv*' AND 'intervention*') indicating quality improvement in general, as well as the following CQI-related terms: 'Plan-Do-Study-Act', 'Plan-Do-Check-Act', 'Define-Measure-Analyse-Improve-Control', 'Define-Measure-Analyse-Design-Verify', 'iterative cycle', Deming, Taguchi, Kansei, Kaizen, 'six-sigma', 'total quality management','quality function deployment', 'House of quality', 'quality circle', 'quality circles', 'Toyota production system', 'lean manufacturing' and 'business process reengineering'. The search resulted in 9427 articles.

To identify candidate QII articles from this set, two authors (LR, PS) used previously described definitions[22] to identify 201 potentially relevant titles and abstracts reporting empirical data on a QII from among 1600 randomly selected articles. We then screened the remainder of the 9427 articles using an experimental machine learning algorithm that utilised the manual title/abstract review as a learning set. We added 49 machine-screened articles that screened in at a maximal confidence level. Finally, we added 24 articles recommended by expert panel members as QII examplars, resulting in a total of 272 candidates.

We identified QII articles from among these 272 using the QII screening form with the explicit criteria discussed above (QII-1 through QII-4).[21] Two reviewers

Table 2 Continuous quality improvement (CQI) features assessment for articles identified as studies of quality improvement interventions (QIIs)

CQI-1	**ITERATIVE DEVELOPMENT PROCESS: Did the improvement initiative involve iterative design AND implementation of a set of specific changes for improving care (ie, a change package)?** *Iterative Development: Cyclical process, such as Plan-Do-Study-Act cycles, in which the initial design and implementation of the set of changes for improving care is followed by redesign and reimplementation.* ***A single iterative cycle thus includes initial implementation followed by assessment, redesign, and reimplementation.*** 1) 0 cycles 2) 1 or unclear # (>0) of complete cycles 3) 2 or more complete cycles
CQI-2	**FEEDBACK AT MEETINGS INVOLVING PARTICIPANT LEADERS: Did leaders of the improvement initiative (eg, local managers, clinical leaders, central experts, or improvement teams) from participating study organisation(s) or local study site(s) meet to review information on initiative implementation?** *Information on Implementation Progress: Includes formal feedback, review of interim outcomes, and/or informal discussions relating to progress on the introduction of a set of changes for improving care (ie, the change package) into organisation(s) or site(s).* *Feedback Meetings Involving Participant Leaders: Can be by telephone or in person, but not by paper or e-mail only (ie, must provide opportunity for interaction). Must include organisation or site leaders and not researchers alone.* 1) No / Don't Know 2) Participant leader meetings, but unclear if improvement initiative implementation discussed 3) Yes, participant leader meetings where improvement initiative implementation discussed
CQI-3	**FEEDBACK OF SYSTEMATICALLY-COLLECTED DATA: Did the improvement initiative include feedback of systematically-collected data on implementation?** *Systematically-collected data: Quantitative or qualitative data, collected according to a design or plan or for which methods are specified in the article. Exclude information produced at a meeting at which random individuals discuss problems. Include only data collected during implementation of a set of changes for improving care (ie, change package).* 1) No / Don't Know 2) Feedback of systemically-collected data on implementation—only one data point 3) Yes, feedback of systematically-collected data on implementation—multiple data points
CQI-4	**RECOGNIZED CHANGE METHOD: Were one or more recognised change methods used in the improvement initiative?** *System change methods such as the following: 'CQI, Continuous Quality Improvement'; DMAIC, Define-Measure-Analyse-Improve-Control; DMADV. Define-Measure-Analyse-Design-Verify; approaches of Deming,Taguchi, Kaizen, Juran, or Kansei; 'six-sigma'; 'total quality management'; 'quality function deployment'; 'House of quality'; 'quality circle'; 'Toyota production system'; 'lean manufacturing'; 'business process reengineering'; CRM, 'crew resource management'; 'Breakthrough Series'; 'Institute for Healthcare Improvement' quality improvement; Evidence-based Quality Improvement.* **Specify other terms used: _____** 1) No / Don't Know 2) Change method mentioned but not explicitly described 3) Yes, specific elements of the change method explicitly described
CQI-5	**DATA-DRIVEN: In your judgment, to what extent was the design AND/OR implementation of a set of changes for improving care (ie, change package) driven by data collected systematically during implementation?** *Local conditions: Barriers, resources, or baseline characteristics of the organisation(s) or local site(s) that could influence outcomes or progress (eg, costs, capabilities, leadership, QI experience; patient population; or provider characteristics).* Very Little or Not At All (1) (2) Somewhat (3) (4) Substantially (5)
CQI-6	**LOCAL CONDITIONS: In your judgment, to what extent were local conditions at study organisation(s) or site(s) taken into account in the design AND/OR implementation of the set of specific changes for improving care (ie, the change package)?** Very Little or Not At All (1) (2) Somewhat (3) (4) Substantially (5)

(MD plus SH or SO) reviewed the full text of each candidate article to apply the screen, and consulted LR for resolution when there was disagreement.

Assessment of CQI features

Two reviewers (YL, RF) pilot tested the initial CQI features assessment on a subset of 45 included QIIs. Two reviewers (YL, SO) applied the final CQI features assessment to the remaining 106 included QIIs.

Analysis

In calculating consensus results, we adjusted for reviewer effect. Some reviewers consistently rate items lower on a scale (ie, the mean, or midpoint, around which their ratings vary is lower) and some reviewers rate consistently higher.[30] Reviewer effect adjustment normalises raters to a common mean. We computed the inter-rater reliabilities of the QII and CQI features assessment using κ statistics (for bivariate assessments) and intra-class correlations (for scales).

We counted a CQI feature 'present' in an article if both reviewers rated that feature as ≥2 (on a three-point scale) or ≥3 (on a five-point scale). We weighted items equally and did not prespecify a cut-off for qualifying a study as 'CQI.' However, to explore potential cut-off points, we created a composite rating by averaging across all CQI features for each article. We applied cut-offs by using the average composite rating across both reviewers, as well as by requiring both reviewers' composite ratings to independently surpass the cut-off. For composite ratings, we analysed results both with and without items CQI-5 and CQI-6 to account for the use of a five-point scale.

RESULTS

QII screen results

QII screening resulted in 151 included QII articles. Inter-rater per cent agreement for application of the explicit screening form (prior to resolution of disagreements) was 85.7% (κ=0.71). The final inclusion set comprised 106 QIIs. Table 3 shows that most reported QIIs were hospital or outpatient based (56% and 33% respectively). Most studies (77%) reported no comparison group and 83% reported improvements following interventions. About half of the articles involved an author who had a PhD or master's degree; 10% indicated an academic professorial type position. Articles appeared predominantly in clinical journals (64%).

Table 3 Descriptive characteristics of quality improvement interventions (QIIs)

Characteristic	QIIs (n=106) n (%)
Setting	
Hospital	59 (56)
Outpatient	35 (33)
Long-term care	9 (9)
Other	3 (3)
Don't know	0 (0)
Evaluation target	
Change package	101 (95)
Change method	2 (2)
Both	1 (1)
Other	0 (0)
Don't know	2 (2)
Evaluation design	
No comparison group/don't know	82 (77)
Randomly assigned comparison group	12 (11)
Non-randomly assigned comparison group	12 (11)
Researcher involvement in authorship	
Professor	11 (10)
PhD	26 (25)
Master's trained	26 (25)
Other	0 (0)
No/don't know	43 (41)
Results	
Reported as showing improvement	88 (83)
Reported as equivocal	11 (10)
Reported as NOT showing improvement	6 (6)
No/don't know	1 (1)
Journal type	
Quality improvement/health services research	38 (36)
Clinical	68 (64)
General	10 (9)
Nursing	17 (16)
Other specialty	41 (39)

CQI features assessment

Table 4 shows inter-rater reliability (intra-class correlation) and per cent agreement between reviewers for CQI features. Per cent agreement ranged from 55.7% to 75.5% for the six items, and reviewer-adjusted intra-class correlations ranged from 0.43 to 0.62 (in the 'fair to good' reliability range).

Among features, feedback of systematically collected data was the most common (64%), followed by being at least 'somewhat' adapted to local conditions (61%), feedback at meetings involving participant leaders (46%), using an iterative development process (40%), being at least 'somewhat' data driven (34%), and using a recognised change method (28%). Articles in quality improvement or health services research journals reported all CQI features more often than clinical journals, significantly more for two features, feedback

Table 4 Continuous quality improvement (CQI) features assessment reliability and feature prevalence

Item	Feature	Feature reliability		Feature prevalence*			
		n=106 articles		All journals (n=106 articles)	Clinical journals (n=68 articles)	QI/HSR journals (n=38 articles)	
		% agreement	ICC† 95% CI	n (%)	n (%)	n (%)	p
CQI-Q1	Iterative development process	59.4	0.43 (0.27 to 0.58)	42 (40)	24 (35)	18 (47)	0.223
CQI-Q2	Feedback at meetings involving participant leaders	58.5	0.54 (0.41 to 0.68)	49 (46)	28 (41)	21 (55)	0.163
CQI-Q3	Feedback of systematically collected data	59.4	0.45 (0.30 to 0.61)	68 (64)	37 (54)	31 (82)	0.005¶
CQI-Q4	Recognised change method	75.5	0.62 (0.50 to 0.74)	30 (28)	15 (22)	15 (40)	0.056
CQI-Q5	Data driven‡	59.4	0.50 (0.36 to 0.64)	36 (34)	17 (25)	19 (50)	0.009§
CQI-Q6	Local conditions‡	55.7	0.52 (0.38 to 0.66)	65 (61)	37 (54)	28 (74)	0.051

*Features were considered 'present' if both reviewers rated the item ≥2.
†Intra-class correlation is reviewer adjusted.
‡Response categories for items CQI-5 and CQI-6 were collapsed to a three-point scale.
§χ^2 Group difference between clinical and quality improvement (QI)/Health Services Research (HSR) categories statistically significant at p=0.05 level.
¶χ^2 Group difference between clinical and QI/HSR categories statistically significant at p=0.005 level.

of systematically collected data and being at least 'somewhat' data driven.

Table 5 shows that 14% of articles included all six CQI features at a score of two or more (see table 2 for scoring). Table 6 shows another approach to assessing cut-offs for considering an article to represent CQI methods. This approach uses a composite rating of 'CQI-ness' based on the average score across all features for each article. Based on achieving a composite score of two or more, 44% of QII articles showed some level of CQI-ness, and could be so identified with a κ for reliability of 0.49 (fair reliability). Depending on the cut-off value used, the number of interventions in our QII sample qualifying as CQI interventions ranged from 1% to 44%.

DISCUSSION

This project used expert consensus methods to develop and apply potential CQI definitional features to a comprehensive sample of QII literature. We found reasonable inter-rater reliability for applying consensus-based features to electronically identified candidate QII articles. This indicates that systematic sample identification of CQI intervention articles is feasible. We found considerable variation in the reporting of individual features.

We aimed to assess the feasibility of creating a consensus-based definition of CQI for evidence review. We found that while experts could agree on a core set of important features, and these features could be reliably applied to literature, few articles contained a consistent core set. Alternatively, we tested a composite measure of 'CQI-ness' that reflected the quantity of CQI features reported. We found that this approach was feasible and may be useful for review purposes. This approach has important limitations, however, in that specific features may be of varying relevance depending on the purpose of the review.

As an illustration of the diversity of articles with CQI features, only one article was maximally rated by both reviewers on all features. Nowhere in that article does the phrase 'CQI' or even 'quality improvement' appear, which shows the disjunct between reporting of CQI features and use of the term 'CQI' itself.

During review, we noted that QII articles were inconsistently organised, with important methodological information about the intervention scattered throughout the sections of the articles. For iterative processes and data feedback in particular (CQI-1, CQI-2, and CQI-3), reviewers often had to extract data from tables (eg, monthly infection rates) rather than the main text. Development of a standard order for reporting CQI methods and results might make CQI articles easier to write and review.

Table 5 Quality improvement intervention (QII) articles, stratified by the number of continuous quality improvement (CQI) features present

No. of CQI features present* (of features 1–6)	QII articles (n=106) Cut-off: feature ratings ≥2 n (%)	Cut-off: feature ratings=3 n (%)
0 of 6 features	15 (14)	47 (44)
1 of 6 features	24 (23)	21 (20)
2 of 6 features	15 (14)	16 (15)
3 of 6 features	14 (13)	10 (9)
4 of 6 features	11 (10)	6 (6)
5 of 6 features	12 (11)	5 (5)
6 of 6 features	15 (14)	1 (1)
No. of CQI features present* (of features 1–4)		
0 of 4 features	28 (26)	50 (47)
1 of 4 features	21 (20)	26 (25)
2 of 4 features	23 (22)	21 (20)
3 of 4 features	14 (13)	7 (7)
4 of 4 features	20 (19)	2 (2)

*'Present' implies both reviewer ratings were greater than or equal to the indicated cut-off. For items CQI-5 and CQI-6, ratings were collapsed to a three-point scale (from the original five-point scale).

Two items, data-drivenness (CQI-5) and degree of adaptation to local conditions (CQI-6), required implicit reviewer judgement due to our inability to develop reliable explicit criteria for assessing them. Some articles, for example, implied data-drivenness by alluding to quantitative audit/feedback mechanisms employed during implementation, but did not display any data. Multisite trials of standardised change packages, as another example, might imply methods for local involvement, but describe local adaptations only vaguely.

An earlier CQI evidence review[7] also identified the issue of variable language use and reporting. Efforts to standardise reporting for randomised controlled

Table 6 Quality improvement interventions articles, stratified by composite rating over all continuous quality improvement (CQI) features

Article composite rating cut-off‡	No. of articles: Average composite rating above cut-off* n (%)	Independent composite ratings above cut-off† n (%)	κ
CQI-1 through CQI-6			
≥2.00	47 (44)	37 (35)	0.49
≥2.25	37 (35)	25 (24)	0.59
≥2.50	23 (22)	19 (18)	0.47
≥2.75	11 (10)	7 (7)	0.38
=3.00	1 (1)	1 (1)	0.11
CQI-1 through CQI-4 only			
≥2.00	47 (44)	40 (38)	0.44
≥2.25	41 (39)	30 (28)	0.58
≥2.50	25 (24)	17 (16)	0.45
≥2.75	10 (9)	10 (9)	0.41
= 3.00	2 (2)	2 (2)	0.21

*Calculated for each article by taking the average of both reviewers' ratings for each item, and then taking the average over all items. For an article to count, the average composite rating had to surpass the indicated cut-off value.
†Calculated for each article, separately for each reviewer, by taking the average rating over all items. For an article to count, both reviewers' independent composite ratings had to surpass the indicated cut-off value.
‡Composite ratings could range from 1.00 to 3.00. For items CQI-5 and CQI-6, ratings were collapsed to a three-point scale (from the original five-point scale).

trials[13–15] and QIIs[31] have proven useful. Our results support similar efforts for CQI interventions.

This study has limitations. The lack of relevant medical subject heading terms for either QII or CQI, in addition to inherent variation in CQI language use, may have reduced search sensitivity. To address this limitation, we used an inclusive electronic search strategy (Hempel *et al*, submitted) and additional expert referral of articles. This in turn resulted in a large candidate article set that required substantial screening. The number of electronically generated articles, however, is within the range of major evidence reviews.[32–34] We further expect that studies may most likely apply our methods to smaller sets addressing CQI subtopics, such as CQI for diabetes. The expert panel portion of this study is limited by involvement of a small though diverse group of key stakeholders. The purpose of the study, however, was to clarify and describe variations in reporting of key CQI features rather than to propose a final definition.

Currently, given the low agreement on the meaning of the term 'CQI', readers can have very little confidence that reviews of CQI interventions will include coherent samples of the literature. Without explicit identification of specific CQI features, reviews will yield uninterpretable results. Continued work assessing CQI features in relevant literature will result in more efficient, effective learning about this important quality improvement approach. Meanwhile, the more explicit CQI authors can be in describing the key features of their CQI interventions,[31] the more interpretable and useful the results of their work will be.

Acknowledgements The authors would like to thank the members of the CQI stakeholder panel: David Atkins, MD, MPH, Department of Veterans Affairs; Frank Davidoff, MD, Institute for Healthcare Improvement; Martin Eccles, MD, Newcastle University (UK), Co-Editor-in-Chief, *Implementation Science*; Robert Lloyd, PhD, Institute for Healthcare Improvement; Vin McLoughlin, PhD, The Health Foundation (UK); Shirley Moore, RN, PhD, Case Western Reserve University; Drummond Rennie, MD, University of California, San Francisco, Deputy Editor, *Journal of the American Medical Association*; Susanne Salem-Schatz, ScD, HealthCare Quality Initiatives; David P. Stevens, MD, Dartmouth University, Editor Emeritus, *BMJ Quality & Safety*; Edward H. Wagner, MD, MPH, Group Health Cooperative; Brian Mittman, PhD, Department of Veterans Affairs, Greater Los Angeles Healthcare System, Co-Editor-in-Chief, *Implementation Science*; and Greg Ogrinc, MD, Dartmouth University. They would also like to thank the following individuals, who helped support and guide this project: Breanne Johnsen, Aneesa Motala, Siddartha Dalal, Kanaka Shetty, and Roberta Shanman, RAND; and Mary Haines, Sax Institute (Australia).

Funding Robert Wood Johnson Foundation (RWJF, PO Box 2316, Route 1 and College Road East, Princeton, NJ 08543, USA) grants were given to LR (Grant ID 65113: Advancing the science of continuous quality improvement: A framework for identifying, classifying and evaluating continuous quality improvement studies and Grant ID 67890: Providing a framework for the identification, classification, and evaluation of quality improvement initiatives). RWJF's assistance in collecting data was limited to allowing the authors to glean definitions of continuous quality improvement from grant applications to the Advancing the Science of Continuous Quality Improvement program. RWJF played no role in the study design, in the analysis and interpretation of data, in the writing of the report, or in the decision to submit the paper for publication. Additional funding was provided by the Agency for Healthcare Research and Quality, the Veterans Health Administration, the British Healthcare Foundation and RAND Health.

Competing interests None to declare.

Ethics approval This study was conducted with the approval of the RAND Corporation Human Subjects Protection Committee.

Contributors LR, PS, SH, and MD conceived the study design. LR, PS, RF, YL, SH, SO, and MD were involved in data collection. All authors were involved in the analysis and interpretation of data. SO drafted the article; all authors were involved in critical revisions and approved the final version.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Berwick DM. Continuous improvement as an ideal in health care. *N Engl J Med* 1989;320:53–6.
2. Laffel G, Blumenthal D. The case for using industrial quality management science in health care organizations. *JAMA* 1989;262:2869–73.
3. Berwick DM, Godfrey AB, Roessner J. *Curing health care*. San Francisco: Jossey-Bass, 1990.
4. Institute for Healthcare Improvement. *Model for improvement*. http://www.ihi.org/IHI/Topics/Improvement/ImprovementMethods/HowToImprove/ (accessed 1 Dec 2010).
5. Deming WE. *Out of the crisis*. Cambridge: MIT Press, 1986.
6. Juran JM. *Juran on planning for quality*. New York: Free Press, 1988.
7. Shortell SM, Bennett CL, Byck GR. Assessing the impact of continuous quality improvement on clinical practice: what it will take to accelerate progress. *Milbank Q* 1998;76:593–624.
8. Flyvbjerg B. *Making social science matter: why social inquiry fails and how it can succeed again*. Cambridge: Cambridge University Press, 2001.
9. Manville B. Knowledge management and complex adaptive systems: the case of McKinsey & Co. In. Clippinger J, ed. *The biology of business*. New York: Jossey-Bass, 1999.
10. Krieger JN, Nyberg L Jr, Nickel JC. NIH consensus definition and classification of prostatitis. *JAMA* 1999;282:236–7.
11. Bingham CO 3rd, Pohl C, Woodworth TG, *et al*. Developing a standardized definition for disease 'flare' in rheumatoid arthritis (OMERACT 9 Special Interest Group). *J Rheumatol* 2009;36:2335–41.
12. Morse RM, Flavin DK. The definition of alcoholism. The Joint Committee of the National Council on Alcoholism and Drug Dependence and the American Society of Addiction Medicine to study the definition and criteria for the diagnosis of alcoholism. *JAMA* 1992;268:1012–14.
13. Begg C, Cho M, Eastwood S, *et al*. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637–9.
14. Boutron I, Moher D, Altman DG, *et al*. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295–309.
15. Moher D, Schulz KF, Altman DG; CONSORT GROUP (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;134:657–62.
16. Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem* 2003;49:1–6.
17. Von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573–7.
18. Moher D, Cook DJ, Eastwood S, *et al*. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999;354:1896–900.
19. Stroup DF, Berlin JA, Morton SC, *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008–12.
20. Robert Wood Johnson Foundation. *Advancing the science of continuous quality improvement*. http://www.rwjf.org/applications/solicited/cfp.jsp?ID=20161 (accessed 14 Dec 2010).
21. Danz MS, Rubenstein LV, Hempel S, *et al*. Identifying quality improvement intervention evaluations: is consensus achievable? *Qual Saf Health Care* 2010;19:279–83.

22. 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.
23. McLaughlin C, Kaluzny A. *Continuous quality improvement in health care: theory, implementation, and applications*. London, UK: Jones and Bartlett Publishers International, 2004.
24. Davidoff F, Batalden PB. What is 'quality improvement' and how can it transform healthcare? *Qual Saf Health Care* 2007;16:2–3.
25. Lynn J, Baily MA, Bottrell M, *et al*. The ethics of using quality improvement methods in health care. *Ann Intern Med* 2007;146:666–73.
26. *Cochrane effective practice and organisation of care group*. http://epoc.cochrane.org/ (accessed 14 Dec 2010).
27. Mathew JP, Glas K, Troianos CA, *et al*. ASE/SCA recommendations and guidelines for continuous quality improvement in perioperative echocardiography. *Anesth Analg* 2006;103:1416–25.
28. Coleman MT, Endsley S. Quality improvement: first steps. *Fam Pract Manag* 1999;6:23–6.
29. MRC Health Services and Public Health Research Board. *A framework for development and evaluation of RCTs for complex interventions to improve health*. Medical Research Council, 2000. http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003372 (accessed 5 Oct 2010).
30. Rubenstein LV, Kahn KL, Harrison ER, *et al*. *Structured implicit review of the medical record: a method for measuring the quality of inhospital medical care and a summary of quality changes following implementation of the Medicare prospective payment system*. Santa Monica: RAND Corporation, 1991:19.
31. Davidoff F, Batalden P, Stevens D, *et al*; Standards for QUality Improvement Reporting Excellence Development Group. Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *J Gen Intern Med* 2008;23:2125–30.
32. Shekelle PG, Morton SC, Keeler EB. Costs and benefits of health information technology. *Evid Rep Technol Assess (Full Rep)* 2006;132:1–71.
33. Fung CH, Lim YW, Mattke S, *et al*. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med* 2008;148:111–23.
34. Hussey PS, de Vries H, Romley J, *et al*. A systematic review of health care efficiency measures. *Health Serv Res* 2009;44:784–805.