


RESEARCH ARTICLE

Open Access



# Approaches of integrating the development of guidelines and quality indicators: a systematic review

Miranda W. Langendam<sup>1</sup>, Thomas Piggott<sup>2</sup>, Monika Nothacker<sup>3</sup>, Arnav Agarwal<sup>4</sup>, David Armstrong<sup>5</sup>, Tejan Baldeh<sup>1</sup>, Jeffrey Braithwaite<sup>6</sup>, Carolina Castro Martins<sup>2,7</sup>, Andrea Darzi<sup>2</sup>, Itziar Etxeandia<sup>8</sup>, Ivan Florez<sup>1,9</sup>, Jan Hoving<sup>10</sup>, Samer G. Karam<sup>2</sup>, Thomas Kötter<sup>11</sup>, Joerg J. Meerpohl<sup>12</sup>, Reem A. Mustafa<sup>2,13</sup>, Giovanna E. U. Muti-Schünemann<sup>14</sup>, Philip J. van der Wees<sup>15</sup>, Markus Follmann<sup>16</sup> and Holger J. Schünemann<sup>2,17,18\*</sup> 

## Abstract

**Background:** Guidelines and quality indicators (for example as part of a quality assurance scheme) aim to improve health care delivery and health outcomes. Ideally, the development of quality indicators should be grounded in evidence-based, trustworthy guideline recommendations. However, anecdotally, guidelines and quality assurance schemes are developed independently, by different groups of experts who employ different methodologies. We conducted an extension and update of a previous systematic review to identify, describe and evaluate approaches to the integrated development of guidelines and related quality indicators.

**Methods:** On May 24th, 2019 we searched in Medline, Embase and CINAHL and included studies if they reported a methodological approach to guideline-based quality indicator development and were published in English, French, or German.

**Results:** Out of 16,034 identified records, we included 17 articles that described a method to integrate guideline recommendations development and quality indicator development. Added to the 13 method articles from original systematic review we included a total 30 method articles. We did not find any evaluation studies. In most approaches, guidelines were a source of evidence to inform the quality indicator development. The criteria to select recommendations (e.g. level of evidence or strength of the recommendation) and to generate, select and assess quality indicators varied widely. We found methodological approaches that linked guidelines and quality indicator development explicitly, however none of the articles reported a conceptual framework that fully integrated quality indicator development into the guideline process or where quality indicator development was part of the question formulation for developing the guideline recommendations.

**Conclusions:** In our systematic review we found approaches which explicitly linked guidelines with quality indicator development, nevertheless none of the articles reported a comprehensive and well-defined conceptual framework which integrated quality indicator development fully into the guideline development process.

**Keywords:** Guidelines, Recommendations, Quality improvement, Quality assurance

\* Correspondence: [schuneh@mcmaster.ca](mailto:schuneh@mcmaster.ca)

<sup>2</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada

<sup>17</sup>Department of Medicine, Hamilton, McMaster University, Hamilton, Canada

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Contributions to the literature

- Quality indicators are used to monitor guideline adherence as they measure structures, processes and health outcomes of care.
- Ideally, development of the quality indicators should be integrated in the guideline development process to establish a direct link with the recommendations.
- We extended and updated a systematic review on existing approaches for integrated development and found that quality indicators development is a topic of high interest, but there is minimal methodological advancement and the connection with guideline development methods is very limited.
- A well-defined methodological framework to integrate quality indicator development fully into the guideline development process is needed.

## Introduction

Guidelines and quality assurance (QA) schemes both aim to improve health care delivery and health outcomes. A QA scheme is a common set of quality and safety requirements for health care service. It covers interventions and services and may include several quality dimensions. Quality indicators are used to benchmark the fulfilment of a requirement using a clearly defined numerator and denominator (ISO 9000:2015 Quality management systems - Fundamentals and vocabulary ([www.iso.org/standard/45481.html](http://www.iso.org/standard/45481.html))). Quality indicators are measurable items referring to structures, processes, and outcomes of care [1]. Ideally, the development of quality indicators should be grounded in evidence-based health care recommendations, derived from trustworthy guidelines.

However, anecdotally, guidelines and quality assurance schemes are developed separately, in isolation, by different groups of experts who employ different methodologies. It is often unclear, for example, how QA organizations (e.g. International Society for Quality in Healthcare) and guideline developers (e.g. the World Health Organization or professional societies) interact and how, when and in which context guideline recommendations are used to develop QA schemes or quality indicators. This lack of coherence may have important adverse consequences for implementation and adherence with respect to guidelines and QA schemes.

There is potential benefit of aligning activities and methods, resulting in an integrated approach. In this context integration means that QA scheme or quality indicator development is considered in all steps of guideline recommendation development, starting with the formulation of the key questions and defining the outcomes. Integration will then result in a set of quality indicators (and QA scheme) that is directly related to the

key questions and recommendations in the guideline. The European Commission (EC) in its European Commission Initiative on Colorectal Cancer (ECICC) is exploring ways to integrate guideline recommendation and QI development.

To inform the ECICC, we performed a systematic review in order to identify and evaluate the current approaches to guideline-based quality indicator development, which is presented in this paper. We also performed a feasibility study, creating a proposal for uniform definitions of QI, performance measures and performance indicators, and we organized a three-day expert workshop, the outcomes of which are published elsewhere (refs Terminology and TwoWorlds paper).

The objectives of this systematic review were twofold. First, to identify and describe approaches that are utilized to develop guideline recommendations and quality indicators, i.e. in an integrated framework. Second, to evaluate the effects of an integrated guideline and quality indicator development approach on individual health outcomes as well as process and structure outcomes (e.g. time required to develop recommendations and quality indicator, feasibility, acceptability by key stakeholders, and development costs).

## Methods

We initially performed a systematic review of peer-reviewed and grey literature to identify approaches in which guideline recommendations and QI are developed in an integrated framework. The protocol with detailed methods description is published in the Prospective Register of Systematic Reviews (PROSPERO, CRD42018097302). We then identified a published systematic review on this topic that was current until April 2010 [2]. We contacted the lead author who agreed to collaborate on an update of that review by first applying the original search strategy and eligibility criteria and expanding it by searching for additional articles and reports.

## Data sources and searches

The eligibility criteria of the original review were English, French, or German articles reporting at least one methodological approach to guideline-based quality indicator development. We searched in Medline, Embase and CINAHL. All study and publication types were included. Studies at the full-text screening stage that did not describe the extraction of recommendations from clinical guidelines in detail were excluded [2].

We refined the eligibility criteria. Reports describing or evaluating approaches in which QA schemes or quality indicators are developed simultaneously or integrated with health-related evidence-based guideline recommendations were eligible for inclusion. Development of

guideline recommendations should be based on evidence summaries, while quality indicator development could be based on evidence or expert consensus, or a combination of the two.

We updated the search as described by the original review by Kötter et al., from 2010 to May 24th, 2019 (See Additional file 1: Appendix A for details on search strategy). In addition, we actively searched for manuals or methods articles that apply guideline-based methods but do not describe that method in detail (topic articles) and we expanded the list of institutional websites (Additional file 1: Appendix B).

### Study selection

Two reviewers independently selected potentially eligible articles by screening titles and abstracts followed by full-text screening (Additional file 1: Appendix C). Disagreements were resolved by discussion, or with the help of a third reviewer.

### Data extraction and quality assessment

Characteristics of the approaches meeting inclusion criteria were abstracted into a standardized form (Additional file 1: Appendix D) [2]. Data were extracted by one reviewer and checked for accuracy by a second reviewer.

For evaluation studies we planned to evaluate the risk of bias of the included studies with a tool appropriate for the study design. However, we did not find any of these studies.

### Data synthesis and analysis

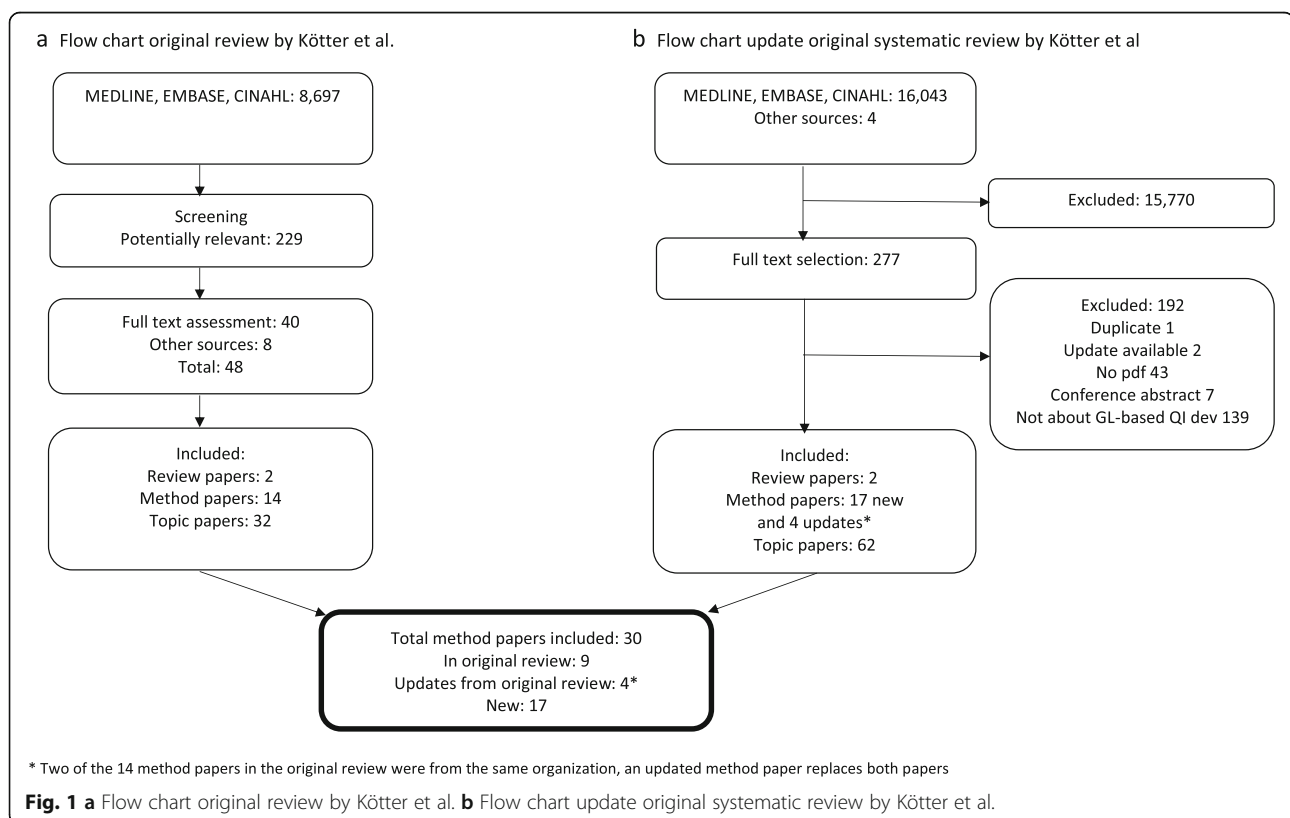
We structured the data in two ways. First, we matched the reported quality indicator process to the guideline development process using the Guidelines International Network (GIN)-McMaster Guideline Checklist (which currently does not include a quality indicator section) [3]. Second, we described the results using the items of the GIN Reporting standards for guideline-based performance measures of GIN [4].

## Results

### Search and selection

Figure 1 presents the flowcharts with the results of the search and selection. The review by Kötter et al. included 48 articles [2]. Of these, 14 articles were method articles, 32 articles were topic articles and 2 were review articles.

In the update (May 24th, 2019) the total number of references from the electronic databases was 16,034 records, of which 273 were screened as potentially relevant. We found an additional 4 articles via other sources (via experts). Of these, 139 articles were not about



guideline-based quality indicator development, 7 were conference abstracts, for 43 we could not retrieve the full text, for 2 articles an update was available and 1 article was a duplicate of another – these 192 articles were excluded. The remaining 85 articles were included: 17 new method articles, 62 topic articles and 2 review articles. Four articles were updates from articles in the original review [5–8]. Two of the 14 method papers in the original review were from the same organization, an updated method paper replaces both papers. Except for the articles with an updated version there was no overlap in included studies between the original review and the update. Thus, in total, 30 method articles were included of which 17 were not included in the prior review [9–25].

The review of institutional websites did not reveal any additional method articles. We did not identify additional method articles among the topic articles and there were no evaluation studies.

#### Characteristics of the included articles

Table 1 presents the study characteristics. The articles were authored by a wide variety of professional societies, universities and governmental organizations across different healthcare settings and clinical topics, based in the United States ( $n = 11$ ), United Kingdom ( $n = 6$ ), Netherlands ( $n = 4$ ), Germany ( $n = 3$ ), Canada ( $n = 2$ ), Belgium ( $n = 1$ ) and Japan ( $n = 1$ ). Two articles were authored by an international group.

#### Approaches that link guideline recommendations and QI development

For a detailed overview we organised the results of the 30 method articles in two ways. The first way is the GIN-McMaster Guideline checklist (Table 2), to match the domains in the guideline process with the accompanying domain in QI development [3]. The second way is the GIN Reporting standards for guideline-based performance measures. This reporting standard includes 9 items of the quality indicator development process [4]. We used these items to provide a detailed overview of the 30 method articles (Table 3).

#### Quality indicator development and GIN-McMaster guideline checklist (Table 2)

Most of the domains in guideline development had a corresponding domain in quality indicator development, but reporting was not optimal. The methods varied in many aspects, for example for group membership (domain 3) and the criteria for quality indicator selection (domain 5). None of the methods papers reported explicitly on establishing group processes, reporting and peer review, conflicts of interest considerations and updating of the quality indicators.

Nine of the 30 method articles described an approach based on one specific guideline [6, 11, 13–15, 21, 26, 32, 37]. While in the other articles, multiple guidelines and other sources were used in order to select potential quality indicators (domain 11).

#### Quality indicator development process (Table 3)

The overall approach observed in quality indicator development was that, based on the guideline(s) and other sources, a list of potential quality indicator was compiled (item 1 and 2 GIN reporting standards). The quality of the guidelines which were used as source for the quality indicator was appraised in 11 method articles, the Appraisal of Guidelines for Research and Evaluation (AGREE) tool was used in 8 of these and in four the criteria for appraisal was not fully specified (item 1b) [6, 7, 9, 14, 17, 23–25, 31, 36, 37]. This list with potential quality indicators served as input for a consensus approach, formal methods, often a modified RAND/UCLA (Delphi) approach, as well as informal methods were used to select the final set of quality indicator (item 3). In the included articles the criteria for selection of quality indicator varied, but could be grouped into relevance, evidence-based, feasibility and measurability (item 4). The potential for quality improvement and improving patient outcomes, scientific soundness and feasibility were most often mentioned as criteria for selecting quality indicator (item 4). How the criteria were defined and scored was not described in detail in most reports (item 5). Quality improvement was the most often mentioned reason for quality indicator development, but in 11 articles this was unclear or not reported (item 6). A practice test was planned for the majority of the methods (item 7). Evaluation of the quality indicator set was largely not reported (item 8). The panel composition was mostly multidisciplinary (item 8), there was no patient involvement in over half of the methods articles (item 9).

#### Linking guideline recommendations to quality indicator development

All but one method article started with describing the quality indicator development process and how the evidence reported in guidelines was used. One article described both development of recommendations and the set of quality indicators for those recommendations. It was unclear, however, how recommendations and quality indicators were linked [15]. None of the articles reported a framework in which quality indicator development was part of the question formulation for developing the guideline recommendations.

Seven of the 30 approaches linked guideline recommendations to quality indicator development, albeit in different ways and for different purposes [11, 14, 15, 21, 26, 32, 37]. Examples of the different purposes are to integrate quantitative measurements of quality and performance into the

**Table 1** General characteristics of the included method papers

Reference	Institution (country)	Topic	Setting	Funding
Advani 2003 [26]	BMIR (US)	Hypertension	–	Public
AHCPR 1995 [27]	AHRQ (US)	–	–	Combined public/private
AHRQ 1995 [28]				
Hughes 2008 [8]				
AQUA 2010 [29]	AQUA (DE)	–	–	Not reported
AQUA 2013 [7]				
ÄZQ 2011 [30]	ÄZQ (DE)	Chronic heart failure	–	Public/private
Nothacker 2011 [6]				
Baker 1995 [31]	Eli Lilly National Clinical Audit Centre (UK)	–	–	Not reported
Bayley 2018 [15]	Institute National d'Excellence en sante et en services sociaux (INESS/ONF) (CA)	Traumatic brain injury	–	Public
Califf 2002 [32]	DCRI (US)	Cardiovascular medicine	All	Public
Cottrell 2018 [24]	American Rhinological Society (US)	Chronic rhinosinusitis	All	Not reported
Cheng 2010 [9]	American Academy of Neurology (US)	Parkinson disease	All	Not reported
Davies 2011 [19]	University Bristol (UK)	Wound, end of life and diabetes care	Outpatient care	Public
Duffy 2005 [33]	APIRE (US)	Bipolar disorder	Outpatient care	Not reported
Fiset 2019 [23]	University of Ottawa School of Nursing (CA)	Pain management	Inpatient, outpatient, long-term care, palliative	Not reported
Follmann 2017 [14]	German Guideline Program in Oncology (GGPO) (DE)	Oncology		Not reported
Golden 2008 [34]	UAMS (US)	Bipolar disorder	Rehabilitation	Public
Graham 2009 [35]	Impact (UK)	–	All	Public
Hommel 2016 [16]	Internal Medicine and IQ healthcare, Radboud UMC (NL)	Perioperative diabetes care	All	Public
Hutchinson 2003 [36]	SCHARR (UK)	CHD	Hospital	Combined public/private
Kahn 2014 [11]	American Thoracic Society (ATS) (US)	Pulmonary, critical care, and sleep medicine	–	Not reported
Laclair 2001 [37]	VA Medical Center (US)	Stroke	Rehabilitation	Public
Luitjes 2013 [25]	Dutch Society for Obstetrics and Gynaecology and IQ healthcare, Radboud UMC (NL)	Hypertensive diseases in pregnancy	Hospital	Public
Mazzone 2014 [10]	American College of Chest Physicians (CHEST)(US)	Lung cancer	District nursing	Private non-profit
Rushforth 2015 [20]	University Leeds (UK)	–	Hospital	Public
Schleedoorn 2016 [13]	EndoKey Group (international)	Endometriosis	Hospital	No funding
Spertus 2005 [38]	American Heart Association (US)	–	Hospital	Public
Spertus 2010 [5]				
Sutcliffe 2012 [18]	NICE (UK)	–	Primary care	Public
Ten Berg 2019 [21]	Dutch Childhood Oncology Group	Pediatric febrile neutropenia	Hospital	Public
Ueda 2019 [22]	Kyoto University, Nara Medical University (Japan)	Low-risk labour care	Hospitals	Public
Vasse 2012 [17]	International	Psychosocial care in dementia	All	Public
Werbrouck 2013 [12]	EFFECT project/KCE (BE)	Uterine cancer	Hospital	Private non-profit
Wollersheim 2007 [39]	IQ healthcare, Radboud UMC (NL)	Oncology, diabetes, antibiotics for pneumonia	–	Not reported



**Table 2** Methods of guideline-based QI development matched to the steps in the guideline development process

Item on GIN-McMaster Guideline checklist	Results in methods papers on QI development
1. Organization, Budget, Planning and Training	Funding: 14 of the 30 approaches were publicly funded, 3 were privately funded, 2 were funded both publicly and privately, 1 did not receive funding and for 10 funding was not reported.
2. Priority Setting	See 5.
3. Guideline Group Membership	Criteria for selection of GDG members were reported in 6 articles. Four articles reported selection of a multidisciplinary panel, including methodological competence, experience in quality improvement, policy decision making and knowledge translation. All 6 articles mentioned clinical expertise for the specific health care topic as competence. Criteria for selection of QI development panel members were mentioned in 15 articles. Clinical expertise was a criterion in all 15 articles, methodological experience was reported in 6 of the 15 articles. Patients/lay persons were part of three panels. Six reports did not use a formal panel and in 9 articles the criteria were unclear.
4. Establishing Guideline Group Processes	Group processes were not described in any article.
5. Identifying Target Audience and Topic Selection	Fifteen articles reported criteria for selecting the QI topics and the target audience. The criteria, and phrasing of the criteria, varied from article to article. Criteria for topic selection included relevance for the specific care domain (e.g. primary care), quality of care gap, sound evidence base, feasibility, availability, measurability, reliability, validity, regulatory requirements, unknown quality adherence, expected impact on quality of life, costs, work load, disease severity, potential to reduce health inequities and covering all aspects of the care process.
6. Consumer and Stakeholder Involvement	Patients were included in the QI selection process in 9 of the 30 articles.
7. Conflict of Interest (COI) Considerations	Conflicts of interest considerations for the QI development process were not mentioned in any of the papers.
8. (PICO) Question Generation	See item 5 (QI topic selection).
9. Considering Importance of Outcomes and Interventions, Values, Preferences and Utilities	Seventeen articles reported criteria for QI selection. In 9 of these articles patient outcomes, health gain or importance or clinical effectiveness were part of the criteria.
10. Deciding what Evidence to Include and Searching for Evidence	See item 11.
11. Summarizing Evidence and Considering Additional Information	All articles used evidence of guidelines as starting point for the QI development (this was an inclusion criterion). Thirteen articles report additional sources, e.g. literature searches for existing QI sets or available data. In 7 articles QI development was based on multiple guidelines, and in 9 articles QI were developed based on one guideline. In 1 article this was not specified.
12. Judging Quality, Strength or Certainty of a Body of Evidence	8 of the 13 articles which report criteria for selecting recommendations as basis for QI development use level of evidence as a criterion; 3 of the 8 approaches use GRADE and suggest that only strong recommendations should be considered for translation into QI.
13. Developing Recommendations and Determining their Strength	See item 12.
14. Wording of Recommendations and of Considerations of Implementation, Feasibility and Equity	Feasibility was mentioned as a criterion for selecting QI (10 articles). Equity was mentioned once, as a criterion for selecting the topic for which the QI were developed.
15. Reporting and Peer Review	Reporting and peer review of QI were not mentioned.
16. Dissemination and Implementation	Implementation: 12 articles report a QI implementation plan as part of their approach, mostly consisting of development of tools and software, and audits.
18. Updating	Updating of QI was not explicitly mentioned in any of the papers.

development cycle of existing and future therapeutics (via guidelines) [32], to derive structured quality indicator and auditing protocols from formalized specifications of guidelines used in decision support systems [26] and to measure quality of care, adherence to guideline recommendations, internal quality management for medical institutions and for benchmarking with other institutions [14, 21].

Three of the seven approaches with a linked approach used the level of evidence to select recommendations suitable for QI development [11, 14, 32]. The method articles from the American Thoracic Society (ATS) and the German Guidelines Program in Oncology (GGPO) report using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach

**Table 3** Guideline-based QI development reporting standard items and report of these criteria in the method papers

Reporting standard item	Reported in method papers
1a. Guideline selection: criteria	<ul style="list-style-type: none"> <li>Selection of guidelines was based on topic and               <ul style="list-style-type: none"> <li>evidence-based development (<math>n = 18</math>)</li> <li>methodological quality of the guideline (<math>n = 2</math>)</li> <li>use of GRADE (<math>n = 1</math>)</li> <li>structured format (<math>n = 1</math>)</li> <li>no other criterion (<math>n = 1</math>)</li> <li>unclear (<math>n = 7</math>)</li> </ul> </li> </ul>
1b. Guideline selection: appraisal of guidelines	<ul style="list-style-type: none"> <li>AGREE (<math>n = 8</math>)</li> <li>criteria not fully specified (<math>n = 4</math>)</li> <li>not reported (<math>n = 18</math>)</li> </ul>
2. Selection of guideline recommendations	<ul style="list-style-type: none"> <li>based on topic (<math>n = 2</math>)</li> <li>impact on patient outcome (<math>n = 4</math>) / burden of illness (<math>n = 1</math>) / clinical utility (<math>n = 1</math>) / available treatment (<math>n = 1</math>)</li> <li>relevance (<math>n = 4</math>) / appropriateness (<math>n = 1</math>)</li> <li>value for money (<math>n = 1</math>)</li> <li>practice variability (<math>n = 1</math>)</li> <li>scope for improvement (<math>n = 1</math>) / gap in quality of care (<math>n = 1</math>)</li> <li>priority / feasibility for implementation (<math>n = 3</math>)</li> <li>validity (<math>n = 2</math>) / reliable (<math>n = 1</math>)</li> <li>(high) level of evidence (<math>n = 8</math>) / adequate scientific proof (<math>n = 1</math>)</li> <li>direct link to aim of guideline (<math>n = 1</math>)</li> <li>common to more than one guideline (<math>n = 1</math>)</li> <li>unclear (<math>n = 3</math>)</li> <li>no selection (<math>n = 6</math>)<sup>a</sup></li> </ul>
3. Selection of performance measures from recommendations	<ul style="list-style-type: none"> <li>formal panel method (<math>n = 11</math>)</li> <li>other or informal consensus method (<math>n = 13</math>)</li> <li>not reported (<math>n = 2</math>)</li> <li>unclear (<math>n = 4</math>)</li> </ul>
4. Core attributes of performance measures (criteria for selecting QI)	<ul style="list-style-type: none"> <li>relevance (<math>n = 4</math>)</li> <li>potential for improvement (<math>n = 9</math>) / likely to change current practice (<math>n = 2</math>) / gap in quality of care (<math>n = 2</math>) / importance for health care (<math>n = 4</math>)</li> <li>burden of illness (<math>n = 2</math>) / improving patient outcomes (<math>n = 9</math>)</li> <li>cost-effectiveness (<math>n = 4</math>)</li> <li>influenced by service provider (<math>n = 3</math>)</li> <li>appropriateness (<math>n = 1</math>)</li> <li>evidence base/scientific soundness (<math>n = 7</math>)</li> <li>(strength of) association with patient important outcome (<math>n = 2</math>)</li> <li>feasibility (<math>n = 7</math>)</li> <li>no risk for unintended consequences (<math>n = 3</math>)</li> <li>unambiguous definition (<math>n = 2</math>) / clear (<math>n = 1</math>)</li> <li>data routinely collected (<math>n = 1</math>)</li> <li>measurable (<math>n = 4</math>) / interpretable (<math>n = 1</math>) / actionable (<math>n = 2</math>)</li> <li>applicable (<math>n = 3</math>) / acceptable (<math>n = 1</math>) / adherence (<math>n = 1</math>)</li> <li>reliable (<math>n = 6</math>) / face validity (<math>n = 2</math>) / construct validity (<math>n = 1</math>) / content validity (<math>n = 2</math>)</li> <li>precision (<math>n = 1</math>)</li> <li>minimum bias (<math>n = 1</math>)</li> <li>not reported (<math>n = 4</math>)</li> <li>unclear (<math>n = 3</math>)<sup>a</sup></li> </ul>
5. Specification of performance measures Numerator and denominator is specified unambiguously and in detail.	<ul style="list-style-type: none"> <li>denominator: population eligible to receive the clinical interventions, numerator: desired intervention and subset of population that should receive it (<math>n = 6</math>)</li> <li>based on algorithm (<math>n = 1</math>)</li> <li>formulation of numerator and denominator in line with formulation of recommendation (<math>n = 1</math>)</li> <li>numerator and denominator including risk adjustment factors (<math>n = 3</math>)</li> <li>clinical researcher drafted an expanded text for each recommendation, using logical operators (e.g. 'AND' and 'OR') to link descriptive statements to produce numerators and denominators (<math>n = 1</math>)</li> <li>method not specified in detail (<math>n = 15</math>)</li> <li>not reported (<math>n = 3</math>)</li> </ul>
6. Intended use of performance measure	<ul style="list-style-type: none"> <li>quality improvement (<math>n = 10</math>)</li> <li>quality of care delivered (<math>n = 2</math>)</li> <li>monitoring compliance with guideline (<math>n = 4</math>)</li> <li>implementation of care (<math>n = 1</math>)</li> <li>clinical audit (<math>n = 1</math>)</li> <li>pay for performance program (<math>n = 1</math>)</li> <li>not specified (<math>n = 8</math>)</li> <li>unclear (<math>n = 1</math>)</li> </ul>

**Table 3** Guideline-based QI development reporting standard items and report of these criteria in the method papers (*Continued*)

Reporting standard item	Reported in method papers
	<ul style="list-style-type: none"> <li>• not reported (<math>n = 2</math>)</li> </ul>
7. Practice test of performance measures	<ul style="list-style-type: none"> <li>• planned (<math>n = 18</math>)</li> <li>• retrospective (<math>n = 2</math>)</li> <li>• implicit (<math>n = 1</math>) / ad hoc (<math>n = 1</math>)</li> <li>• not reported (<math>n = 8</math>)</li> </ul>
8. Review and evaluation of performance measure	<ul style="list-style-type: none"> <li>• plan for evaluation and updating (<math>n = 3</math>)</li> <li>• evaluation including criteria for retiring (<math>n = 1</math>)</li> <li>• mentioned, but not explained in detail (<math>n = 2</math>)</li> <li>• evaluation not reported, often because QI were developed but not yet implemented (<math>n = 24</math>)</li> </ul>
9a. Composition of the panel	<ul style="list-style-type: none"> <li>• monodisciplinary (<math>n = 2</math>)</li> <li>• multidisciplinary (<math>n = 23</math>)</li> <li>• panel composition not reported (<math>n = 5</math>)</li> </ul>
9b. Composition of the panel: patient involvement	<ul style="list-style-type: none"> <li>• yes (<math>n = 10</math>)</li> <li>• no (<math>n = 17</math>)</li> <li>• depends on guideline (<math>n = 1</math>)</li> <li>• not reported (<math>n = 2</math>)</li> </ul>

<sup>a</sup>multiple criteria per methodological framework

to develop recommendations and suggest that strong recommendations should be considered for translation into quality indicators [11, 14]. The third article mentioned level of evidence as an important characteristic of guidelines but was not explicit on how to use it in quality indicator development [32].

Three articles reported challenges for linkage. For Kahn and colleagues, rewording the recommendations to quality indicators and translating the quality indicators into measurable performance indicators with clearly defined numerators and denominators was challenging [11]. In two articles the challenges referred to the use of evidence, or more specifically, the lack thereof. Schlee-doorn and colleagues reported that 11 of the 17 selected recommendations were good practice points (described as expert opinion), and six recommendations were derived from evidence described as Level A. According to the authors, this demonstrates the importance of expert opinion in daily practice [13]. Werbrouck et al. addressed this point as well, where authors remarked that very few process quality indicators in the final list had a high level of evidence. They stated that, “this is either due to the difficulty of providing a high level of evidence for some processes, such as pathology, or due to a real lack of clear evidence from randomized controlled trials for some clinical questions, such as the role of lymphadenectomy. The high mean scores attributed to these quality indicators by the expert’s panel clearly indicate their clinical value emphasizing that evidence should not be the only criterion to select quality indicator since it eliminates indicators deemed relevant by consensus.” [12]

## Discussion

### Summary of findings

We conducted an extension and update of a previous systematic review to identify approaches to the

integrated development of guidelines and related quality indicator. We identified 30 articles describing these approaches, however, in general, these were not based on well-defined conceptual frameworks and lacked full integration of the two areas. Our key findings indicate a lack of coherence between the two fields and heterogeneity in methods. For example, the quality of the guidelines was not assessed in the majority of the articles. This suggests that although quality indicator development is often done on the basis of recommendations by reputable organizations, the suitability and quality of the recommendations may not coincide with the goals of quality indicators. There were no studies that evaluating the impact of guideline integrated quality indicator development on health outcomes. Almost 10 years ago, Kötter et al. came to the same conclusion that there continues to be a lack of impact evaluation of integrated frameworks [2].

The original review included 14 method articles and 32 topic articles; in the update (2010–2019) we found 17 new method articles and twice as many topic papers. This suggests that although quality indicator development is a topic of high interest, there is minimal methodological advancement and the connection with guideline development methods is very limited. The reason for the limited connection is not yet clear, and need to be investigated. However, the fact that guideline developers and quality improvement researchers work in silos is well recognized (ref Two Worlds paper). From this systematic review we have learned some lessons that will influence future practice and quality indicator development.

The key findings that will help with selecting elements of the approaches include the different ways evidence is used in the recommendation selection process for generating quality indicator. Some authors report using



certainty in the evidence (or level of evidence) [5, 15, 22, 32–34], others find using evidence challenging and confuse expert opinion with an interpretation of the evidence [12, 13, 40], and only a few approaches use strength of recommendation [6, 11, 14]. This supports the distinction between certainty of evidence and strength of the recommendation, and guidance how to apply these concepts in quality indicator development.

### Strengths and limitations of this review

Strengths of this review include the systematic approach, the conceptual categorization of the findings according to established tools (GIN-McMaster Checklist and GIN Reporting standards for guideline-based performance measures) [3, 4], and the large amount of new information that we revealed. Potential limitations to the methods of this review are the restriction to three languages and the fact that methods papers are sometimes hard to track. We mitigated these limitations by searching for methods papers on the websites of relevant organisations, consulting experts and checking topic papers for references on the method that was used.

### Implications for practice

For developers in both the guidelines and the QA and quality indicator fields, our review has identified a few existing approaches that may be used to support guideline-based quality indicator development and avoid duplication of effort. In addition, our review highlights that better integration should be sought. For example, quality indicator should be thought of during the formulation of guideline questions to achieve better integration. Finally, our review highlights that those working in quality improvement should distinguish between expert opinion and evidence in the development of quality indicator [40].

### Implications for research

Our review has important implications for research that include the development of a well-defined conceptual framework and testing of that framework.

### Conclusions

Our systematic review of the literature resulted in 30 articles describing approaches for guideline-based QA and quality indicator development. In most approaches, guidelines were used as a source of evidence to inform the QI development. The criteria to select recommendations from the guidelines (e.g. level of evidence or strength of the recommendation) and to generate, select and assess quality indicators varied widely and not all quality indicator development criteria may be addressed in guideline development. We found approaches where guideline and quality indicator development were linked

explicitly, but none of the articles reported a well-defined conceptual framework that properly integrated quality indicator in the guideline process. Research and conceptual development needs to be done in this area and we describe some of these advancements in our accompanying articles (reference to both).

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12913-020-05665-w>.

**Additional file 1.** Appendix A. Search strategies. Appendix B. List of predefined organizations for manual search. Appendix C. Screening forms. Appendix D. Data abstraction forms.

### Abbreviations

EC: European Commission; ECICC: European Commission Initiative on Colorectal Cancer; GIN: Guidelines International Network; GRADE: Grading of Recommendations Assessment, Development and Evaluation; PROSPERO: Prospective Register of Systematic Reviews; QA: Quality Assurance

### Acknowledgements

Elena Parmelli for reviewing the protocol and René Spijker, medical information specialist, for help with running the literature searches.

### Authors' contributions

ML, TP and HS were responsible for the conception of the paper and drafted the first version. AD, IF, CCM, RAM, GEUM, AA, TB, SGK, IE, JH, ML, TP, HS, MN, MF contributed to searching, selection and/or data extraction. ML, TP, HS, AD, IF, CCM, RAM, GEUM, AA, TB, SGK, IE, JH, ML, TP, HS, MN, MF, DA, JB, TK, PW contributed to the interpretation and reporting of the results and provided comments on subsequent versions of the article. All authors read and approved the final manuscript prior submission.

### Funding

The project was partially funded by the European Commission under the umbrella of the European Commission Initiatives on Breast and Colorectal Cancer (ECIBC and ECICC). The funding body was not involved in the design of the study, collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

Not applicable.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health institute, Amsterdam, Netherlands. <sup>2</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada. <sup>3</sup>Institute of Medical Knowledge Management, Association of the Scientific Medical Societies, Berlin, Germany. <sup>4</sup>Department of Medicine, University of Toronto, Toronto, ON, Canada. <sup>5</sup>Farncombe Family Digestive Health Research Institute, McMaster University, Hamilton, Canada. <sup>6</sup>Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Rd, Sydney, Australia. <sup>7</sup>Department of Pediatric Dentistry, Dental School, Federal University of Minas Gerais, Belo Horizonte, Brazil. <sup>8</sup>IKOetxe – Ikerkuntza Osaungintza, Health Research, Gipuzkoa, Irun, Basque Country, Spain. <sup>9</sup>Department of Pediatrics, University of Antioquia, Medellin, Colombia. <sup>10</sup>Coronel Institute of

Occupational Health and Research Center for Insurance Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands. <sup>11</sup>Institute of Family Medicine, University Medical Center Schleswig-Holstein, Campus Lübeck, Lübeck, Germany. <sup>12</sup>Institute for Evidence in Medicine, Medical Center and Faculty of Medicine, University of Freiburg, Freiburg, Germany. <sup>13</sup>Department of Internal Medicine, University of Kansas Medical Center, Kansas, USA. <sup>14</sup>Department of Systemic Pathology, Vita Salute San Raffaele University, Milan, Italy. <sup>15</sup>Department of Rehabilitation and IQ healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands. <sup>16</sup>German Cancer Society, Freiburg, Germany. <sup>17</sup>Department of Medicine, Hamilton, McMaster University, Hamilton, Canada. <sup>18</sup>Department of Health Research Methods, Evidence and Impact, McMaster University Health Sciences Centre, Room 2C16, 1280 Main Street West, Hamilton, ON L8N 4K1, Canada.

Received: 23 April 2020 Accepted: 17 August 2020

Published online: 16 September 2020

## References

- Donabedian A. The quality of care. How can it be assessed? *Jama*. 1988; 260(12):1743–8.
- Kötter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators—a systematic review. *Implement Sci*. 2012;7:21.
- Schunemann HJ, Wiercioch W, Etzeandia I, Falavigna M, Santesso N, Mustafa R, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ*. 2014;186(3):E123–42.
- Nothacker M, Stokes T, Shaw B, Lindsay P, Sipila R, Follmann M, et al. Reporting standards for guideline-based performance measures. *Implement Sci*. 2016;11:6.
- Spertus JA, Bonow RO, Chan P, Diamond GA, Drozda JP Jr, Kaul S, et al. ACCF/AHA new insights into the methodology of performance measurement: a report of the American College of Cardiology Foundation/American Heart Association task force on performance measures. *J Am Coll Cardiol*. 2010;56(21):1767–82.
- Judith Nothacker M, Langer T, Weinbrenner S. Quality indicators for National Disease Management Guidelines using the example of the National Disease Management Guideline for "chronic heart failure". [German]. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2011;105(1):27–37.
- AQUA. *Allgemeine Methoden*. 2013.
- Hughes RG, editor. Patient safety and quality: An evidence-based handbook for nurses. (Prepared with support from the Robert Wood Johnson Foundation). AHRQ Publication No. 08-0043. Rockville: Agency for Healthcare Research and Quality; 2008.
- Cheng EM, Tonn S, Swain-Eng R, Factor SA, Weiner WJ, Bever CT Jr, et al. Quality improvement in neurology: AAN Parkinson disease quality measures: report of the quality measurement and reporting Subcommittee of the American Academy of neurology. *Neurology*. 2010;75(22):2021–7.
- Mazzone PJ, Vachani A, Chang A, Detterbeck F, Cooke D, Howington J, et al. Quality indicators for the evaluation of patients with lung cancer. *CHEST*. 2014;146(3):659–69.
- Kahn JM, Gould MK, Krishnan JA, Wilson KC, Au DH, Cooke CR, et al. An official American thoracic society workshop report: developing performance measures from clinical practice guidelines. *Ann Am Thoracic Soc*. 2014;11(4):S186–95.
- Werbrouck J, Bouche G, de Jonge E, Jacomen G, D'Hondt V, Denys H, et al. Evaluation of the quality of the management of cancer of the corpus uteri—selection of relevant quality indicators and implementation in Belgium. *Gynecol Oncol*. 2013;131(3):512–9.
- Schleedoorn MJ, Vermeulen N, Andersson EA, Maasbach C, Van De Ven-Van Ham MP, Hollander M, et al. Selection of key recommendations for the management of women with endometriosis by an international panel of patients and professionals. *Hum Reprod*. 2016;31(6):1208–18.
- German Guideline Program in Oncology (German Cancer Society, German Cancer Aid, Association of the Scientific Medical Societies). Development of guideline based quality indicators: methodology for the German Guideline Program in Oncology, version 2.0. 2017. Available from: <http://www.leitlinienprogramm.onkologie.de/methodik/informationen-zur-methodik/>.
- Bayley MT, Lamontagne ME, Kua A, Marshall S, Marier-Deschenes P, Allaire AS, et al. Unique features of the INESSS-ONF rehabilitation guidelines for moderate to severe traumatic brain injury: responding to Users' needs. *J Head Trauma Rehabil*. 2018;33(5):296–305.
- Hommel I, van Gurp PJ, Tack CJ, Wollersheim H, Hulscher MEJL. Perioperative diabetes care: development and validation of quality indicators throughout the entire hospital care pathway. *BMJ Qual Saf*. 2016;25(7):525–34.
- Vasse E, Moniz-Cook E, Rikkert MO, Cantegreil I, Charras K, Dorenlot P, et al. The development of quality indicators to improve psychosocial care in dementia. *Int Psychogeriatr*. 2012;24(6):921–30.
- Sutcliffe D, Lester H, Hutton J, Stokes T. NICE and the quality and outcomes framework (QOF) 2009–2011. *Qual Prim Care*. 2012;20(1):47–55.
- Davies P, Wye L, Horrocks S, Salisbury C, Sharp D. Developing quality indicators for community services: the case of district nursing. *Qual Prim Care*. 2011;19(3):155–66.
- Rushforth B, Stokes T, Andrews E, Willis TA, McEachan R, Faulkner S, et al. Developing 'high impact' guideline-based quality indicators for UK primary care: a multi-stage consensus process. *BMC Fam Pract*. 2015;16:156.
- ten Berg S, Loeffen EAH, de Wetering MD, Martens DHJ, Ede CM, Kremer LCM, et al. Development of pediatric oncology supportive care indicators: evaluation of febrile neutropenia care in the north of the Netherlands. *Pediatr Blood Cancer*. 2019;66(2):27504. <https://doi.org/10.1002/psc.27504>.
- Ueda K, Kaso M, Ohtera S, Nakayama T. Updating quality indicators for low-risk labour care in Japan using current clinical practice guidelines: a modified Delphi method. *BMJ Open*. 2019;9(2):e023595.
- Fiset VJ, Davies BL, Graham ID, Gifford W, Woodend K. Developing guideline-based quality indicators: assessing gaps in pain management practice. *Int J Evid Based Healthcare*. 2019;09:09.
- Cottrell J, Yip J, Chan Y, Chin CJ, Damji A, de Almeida JR, et al. Quality indicators for the diagnosis and management of chronic rhinosinusitis. *Int Forum Allergy Rhinol*. 2018;8(12):1369–79.
- Luitjes SH, Wouters MG, Franx A, Bolte AC, de Groot CJ, van Tulder MW, et al. Guideline-based development of quality indicators for hypertensive diseases in pregnancy. *Hypertens*. 2013;32(1):20–31.
- Advani A, Goldstein M, Shahar Y, Musen MA. Developing quality indicators and auditing protocols from formal guideline models: knowledge representation and transformations. *AMIA Annu Symp Proc*. 2003;2003:11–5.
- AHCPR. Designing and implementing guidelines-based performance measures. Agency for Health Care Policy and Research. *Qual Lett Healthc Lead*. 1995;7:21–3.
- AHRQ. Using clinical practice guidelines to evaluate the quality of care – Volume 2: Methods. Rockville: US Dept. of Health and Human Sciences; 1995.
- AQUA-Institut: Allgemeine Methoden im Rahmen der sektorenübergreifenden Qualitätssicherung im Gesundheitswesen nach §137a SGB V. 2010, Version 2.0.
- ÄZQ (Ärztliches Zentrum für Qualität in der Medizin): Qualitätsindikatoren Manual für Autoren Neukirchen: Make a book; 2009.
- Baker R, Fraser RC. Development of review criteria: linking guidelines and assessment of quality. *BMJ*. 1995;311(7001):370–3.
- Califf RM, Peterson ED, Gibbons RJ, Garson A Jr, Brindis RG, Beller GA, et al. Integrating quality into the cycle of therapeutic development. *J Am Coll Cardiol*. 2002;40(11):1895–901.
- Duffy FF, Narrow W, West JC, Fochtmann LJ, Kahn DA, Suppes T, et al. Quality of care measures for the treatment of bipolar disorder. *Psychiatr Q*. 2005;76(3):213–30.
- Golden WE, Hermann RC, Jewell M, Brewster C. Development of evidence-based performance measures for bipolar disorder: overview of methodology. *J Psychiatr Pract*. 2008;14(Suppl 2):18–30.
- Graham WJ. Criterion-based clinical audit in obstetrics: bridging the quality gap? *Best Pract Res Clin Obstet Gynaecol*. 2009;23:375–88.
- Hutchinson A, McIntosh A, Anderson J, Gilbert C, Field R. Developing primary care review criteria from evidence-based guidelines: coronary heart disease as a model. *Br J Gen Pract*. 2003;53(494):690–6.
- LaClair BJ, Reker DM, Duncan PW, Horner RD, Hoening H. Stroke care: a method for measuring compliance with AHCPR guidelines. *Am J Phys Med Rehabil*. 2001;80(3):235–42.
- Spertus JA, Eagle KA, Krumholz HM, Mitchell KR, Normand SL. American College of Cardiology and American Heart Association methodology for the

selection and creation of performance measures for quantifying the quality of cardiovascular care. *Circulation*. 2005;111:1703–12.

39. Wollersheim H, Hermens R, Hulscher M, Braspenning J, Ouwens M, Schouten J, Marres H, Dijkstra R, Grol R. Clinical indicators: development and applications. *Neth J Med*. 2007;65:15–22.
40. Schunemann HJ, Zhang Y, Oxman AD. Expert Evidence in Guidelines G. Distinguishing opinion from evidence in guidelines. *BMJ*. 2019;366:4606.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

