


BMJ Open International experiences in the development and implementation of guideline-based quality indicators: a qualitative study

Monika Nothacker ¹, Marie Bolster,¹ Mirco Steudtner,² Katrin Arnold,³ Stefanie Deckert,³ Monika Becker,⁴ Ina Kopp,¹ Jochen Schmitt³

To cite: Nothacker M, Bolster M, Steudtner M, *et al.* International experiences in the development and implementation of guideline-based quality indicators: a qualitative study. *BMJ Open* 2021;**11**:e039770. doi:10.1136/bmjopen-2020-039770

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-039770>).

Received 25 April 2020

Revised 17 December 2020

Accepted 22 December 2020



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Monika Nothacker; nothacker@awmf.org

ABSTRACT

Objectives Evidence-based clinical guidelines play an important role in healthcare and can be a valuable source for quality indicators (QIs). However, the link between guidelines and QI is often neglected and methodological standards for the development of guideline-based QI are still lacking. The aim of this qualitative study was to get insights into experiences of international authors with developing and implementing guideline-based QI.

Setting We conducted semistructured interviews via phone or skype (September 2017–February 2018) with guideline authors developing guideline-based QI.

Participants 15 interview participants from eight organisations in six European and North American countries.

Methods Organisations were selected using purposive sampling with a maximum variation of healthcare settings. From each organisation a clinician and a methodologist were asked to participate. An interview guide was developed based on the QI development steps according to the 'Reporting standards for guideline-based performance measures' by the Guidelines International Network. Interviews were analysed using qualitative content analysis with deductive and inductive categories.

Results Interviewees deemed a programmatic approach, involvement of representative stakeholders with clinical and methodological knowledge and the connection to existing quality improvement strategies important factors for developing QI parallel to or after guideline development. Methodological training of the developing team and a shared understanding of the QI purpose were further seen conducive. Patient participation and direct patient relevance were inconsistently considered important, whereas a strong evidence base was seen essential. To assess measurement characteristics interviewees favoured piloting, but often missed implementation. Lack of measurability is still experienced a serious limitation, especially for qualitative aspects and individualised care.

Conclusion Our results suggest that developing guideline-based QI can succeed either parallel to or following the guideline process with careful planning and instruction. Strategic partnerships seem key for implementation. Patient participation and relevance, measurement of qualitative aspects and piloting are areas for further development.

Strengths and limitations of this study

- This qualitative study explores facilitating and hindering factors for the whole process of developing and implementing guideline-based quality indicator (QI) including context aspects of the guideline groups position and cooperation possibilities in the respective healthcare setting.
- Following purposive sampling of European and North American guideline institutions in different settings, we interviewed guideline methodologists and clinical guideline authors covering a broad range of expertise.
- We provide some practical solutions to overcome challenges reported by the interviewees.
- However, the number of interviewees and guideline organisations were limited. Supplementary qualitative research with guideline groups is recommended, focusing on structural and methodological insights for realising patient relevant QI and considering qualitative aspects as shared decision making.

Trial registration number German Clinical Trials Registry (DRKS00013006).

BACKGROUND

Evidence-based clinical guidelines play an important role in healthcare and can be a valuable source for quality improvement ideally monitored by guideline-based quality indicators (QIs).^{1–3} An QI can be defined as a 'measurable element of practice performance for which there is evidence or consensus that it can be used to assess the quality, and hence change in the quality, of care provided'.⁴ However, the link between guideline recommendations and QI is often neglected and international methodological standards for the development of guideline-based QI are still lacking.^{5–7} To enable an international comparison and harmonisation of approaches, the Performance Measures

Working Group of the Guidelines International Network (G-I-N) published reporting standards for guideline-based performance measures in 2016.⁸

In the context of a broader research project aiming to guide the development of an evidence and consensus based methodological standard for the development of guideline-based QI in Germany (DFG Project No. 289625106; <https://gepris.dfg.de/gepris/projekt/289625106>),⁹ we here report an international qualitative study. Further subprojects included an analysis of existing German guideline-based QI,¹⁰ qualitative interviews with German authors from guidelines with and without development of guideline-based QI,¹¹ a comparison of international and German guideline-based QI published on the same topic¹² and a systematic compilation of existing assessment instruments for QI. The results of this study will be shared within the respective G-I-N working group with the aim to contribute to an international methodological standard.

A systematic search conducted in preparation of the study up to January 2017 yielded no qualitative publications reflecting on the whole methodological process for developing and implementing guideline-based QI.¹³

Aim

This study sought to contribute to the following research questions:

1. What are the implicit and explicit processes in the development of guideline-based QI?
2. What are the international experiences in the development of guideline-based QI?
3. Which factors hinder or facilitate the development of guideline-based QI from clinical guidelines internationally?

In addition, we explored whether and how guideline-based QI were implemented.

METHODS

We chose a qualitative study design to capture experiences and reflections of clinical and methodological experts who have been involved in the development of guideline-based QI using qualitative content analysis, following Mayring.¹³ A study protocol was developed¹⁴ and the study was registered at the German clinical trials register. Study reporting was chosen according to the Consolidated Criteria for Reporting Qualitative Research developed by Tong *et al*¹⁵ (see online supplemental appendix 1).

Researcher characteristics

The study was mainly carried out by three researchers. The senior researcher (MN) has a clinical and public health background with >10 years of experience in guideline and QI development and is first author of the G-I-N reporting standards. The second researcher (MB) has a social science and public health background with previous experience in qualitative research but not in QI development. The third researcher (MS) has a social science and

healthcare management background (for other authors see the Declarations section).

Sampling of interview participants

We used a broad definition of 'guideline-based' QI, including any QI related to an evidence-based guideline. First, international guideline organisations who had developed guidelines with QI were identified based on guidelines from a systematic review comparing QI from German and corresponding international guidelines¹² and a survey on guideline-based QI development among G-I-N members carried out November 2016. The respective guideline organisations were categorised according to

- ▶ Country.
- ▶ Operating level (national, regional and local).
- ▶ Funding sources (public/private).

Interviewees were selected aiming at purposive maximum variation sampling in relation to guideline development organisations and healthcare settings to reflect a wide range of experiences.^{16 17} For each organisation, the aim was to interview both a clinician and a methodologist, who have worked on guideline-based QI, to consider possible different perspectives. The sample size was planned with 16 on the assumption to interview participants of at least two public and two private organisations acting at the local/regional or at the national level in different countries with the aim to expand the sample according to thematic saturation. Interview participants were identified either via guideline documents or through contact with research coordinators at the respective guideline organisations. Interviewees were contacted by MB via email and followed-up via e-mail and phone, interviews were conducted via phone or skype. MB did not have prior existing relationships with any participant. On first contact, possible interview participants were informed about the aims of the study and were sent a topics overview with the possibility to decline. All interview participants gave written informed consent.

Interview questions

An interview guide was developed by MS, MB and MN (see online supplemental appendix 2). Based on the G-I-N reporting criteria⁸ it comprised a mix of open and a few closed questions focussing on methods, experiences and facilitating/hindering factors in the different stages of QI development from guidelines and the role of the interviewee in QI development and QI implementation. The guide was adjusted after a test interview with an experienced guideline and QI methodologist from the UK and followed for all interviews with slight adaptations to the context and follow-up questions where appropriate. To prepare for the interviews MB reviewed the publicly available information on the selected QI and the development process. Interviews were conducted by MB.

Figure 1: Steps of analysis: coding with iterative categorisation

1. analysis step 1: Line by line coding of interviews
 - a) assignment to main-codes from performance measure reporting criteria [see interview guide appendix 1]:
composition of the development panel, selection of clinical guidelines, extraction of recommendations, development process, measure appraisal, measure specification, intended use of the measure, measure testing/ validating, measure review/re-evaluation + additional main-code: QI implementation
 - b) adding subcodes
 - c) adding new (main-)codes for aspects not covered
2. analysis step 2: summarising analysis: building iterative main categories
summarising main codes into main topics
3. final summary of main topics

Figure 1 Steps of analysis: coding with iterative categorisation. QI, quality indicator.

Patient and public involvement

Patients were not directly involved in this study, but the perception of patient involvement, patient contributions and patient relevance of the resulting QI were aspects addressed in each interview.

Data analysis

Interviews were audio-recorded and transcribed verbatim with pseudonymised analysis using qualitative content analysis. **Figure 1** gives an overview of the different steps: first, the transcribed interviews were read by MB and MN to get familiar with the content. Unclear passages could be resolved by verifying the audio-records. Then, a coding grid was developed with deductive main codes using the G-I-N reporting criteria⁸ and ‘implementation (of guideline-based QI)’ as a framework for analysis. Coding was first done by MB using the qualitative research software MaxQDA.¹⁸ During the coding process the deductive codes were supplemented with inductive subcodes using phrases or little text segments as unit (see **figure 2** as an example of a coding tree). New main codes were assigned for aspects not covered. Coding was reviewed by MN and discussed and agreed between MB and MN. To ensure intercoder reliability, a sample of six

interviews was coded by both researchers and compared for consistency with a resulting high level of concordance. The analysis followed next iterative categorisation according to Neale,¹⁹ which allows for a systematic summary and categorisation of themes directly linked to the original data source. Iterative categorisation was carried out by MB, reviewed by MN and discussed with final agreement. Coding grids were analysed per main categories and summarised including facilitating and hindering factors (see online supplemental appendix 3 as example). The main categories were finally grouped into main themes.

RESULTS

Sample characteristics

Potential participants from 8 organisations of the 11 identified by Becker *et al.*¹² were initially contacted, three did not respond. For each non-responder, an alternative organisation was identified and could be recruited. Fifteen semistructured interviews were carried out with eight methodological and seven clinical experts of four organisations in European and four organisations in North American countries (see **table 1** for characteristic of organisations and interviewees, for one organisation none of the contacted clinicians was available). All interviewees had experience in guideline development. Most of the methodologists and half of the clinicians had previous experience in QI development and were involved directly in QI development whereas some clinicians took a more consulting role. The interviewees from an organisation operating at the local level were not involved in QI development as the indicators selected were taken from a third party. Interviews lasted 25–60 min and were largely variable in the level of details describing the QI development processes.

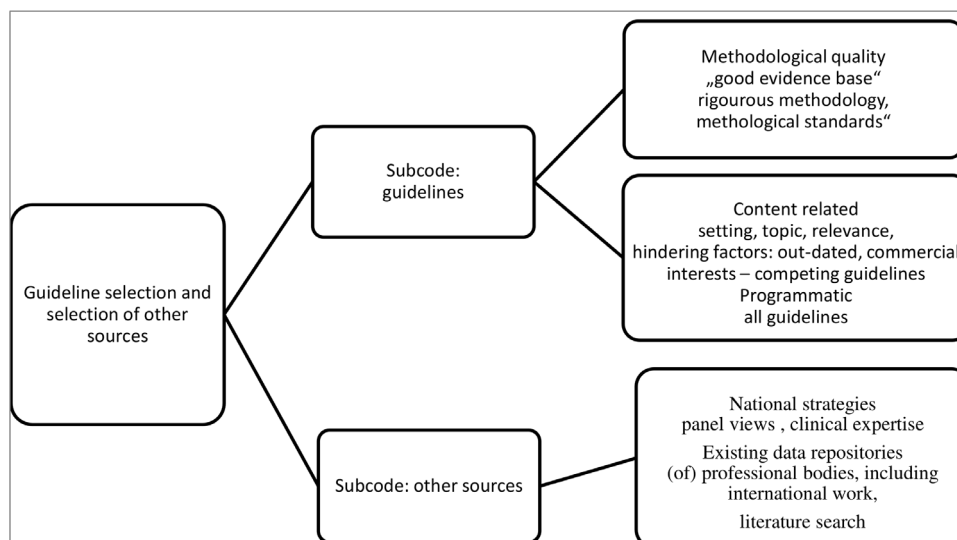

Figure 2 Example of coding tree (analysis step 1).

Table 1 Characteristics of guideline organisations and interviewees

Characteristics of guideline organisations		
Country	Funding	Operational level
European (4): Belgium Scotland Spain UK	Public: 4	National: 4
North American (4): USA (2) Canada (2)	Public: 2 Private: 2	National: 2 Regional: 1 Local: 1
Characteristics of interviewees		
	Methodologists n=8	Clinicians n=7
Current occupation	Full time: 8	Full-time: 2 Part-time: 2 Other (former clinician, clinical coordinator): 3
Experience in guideline development	>1 guideline: 7 1 guideline: 1	>1 guideline: 7
Prior experience with QI development in guidelines apart from current guideline	Yes: 7 No: 1	Yes: 4 No: 3
Role in QI development	None: 1 Involvement ▶ In panel/consensus process only: 0 ▶ In drafting of QI: 7	None: 1 Involvement ▶ In panel/consensus process only: 3 ▶ In drafting of QI: 3

QI, quality indicator.

Processes, experiences and factors hindering or facilitating the development of guideline-based QI

Following iterative categorisation, four main topics were found to be most important for QI development methods from guidelines: (i) organisation/context of guideline and QI development process, (ii) panel composition and decision making, (iii) QI selection criteria/attributes and (iv) intended use and implementation of QI. Table 2 provides an overview with representative quotations and suggestions for approaches according to facilitating factors outlined in the interviews as well as areas for further development. In the following we summarise processes reported as well as experiences of the interview participants for each main topic including facilitating and hindering factors.

1. Organisation/context of guideline and QI development process

The organisation/context of guideline and QI development emerged as one important theme in the interviews leading to an additional main code. Interviewees reported very different ways of organising the development of guideline-based QI. Three organisations developed guideline and QI in one process. The QI development team was identical to the guideline team or consisted

of a subgroup or was led by methodologists with input from the guideline panel. Four organisations developed the guideline first and the QI in a later process with a different team, although individual experts may have been involved in both processes. In one case, a guideline organisation acting at the local level solely adopted existing QI developed by third party groups related to the guideline topic. Finally, one organisation developed informal audit measures with the guideline and national QI were developed in a separate process led by a national group. When guidelines and QI were developed simultaneously, the need to consider QI development early on from the beginning of the guideline process rather than towards the end was considered an important planning factor. To drive QI development forward in a guideline organisation, a dedicated person with QI experience was deemed crucial by several clinicians and methodologists.

Institutional leadership and capacities/resource aspects

A lack of interest and a disjunction between the clinical and political sphere were seen hindering the QI development, whereas strong institutional leadership and capacities were seen favourable. Interviewees from organisations with public and private funding sources stated that resource constraints hampered QI development in general or for certain aspects, for example, piloting. Many concluded that government funding (national or regional) was conducive to QI development. Interviewees judged the QI process to be time-consuming. Time constraints were a hindering factor identified by several interviewees, clinicians and methodologists.

2. Panel composition and decision making

Most interviewees—methodologists and clinicians—stressed the importance of selecting an QI panel that represents different professional groups and settings, is balanced in its representation of interest groups and has a high level of knowledge about QI.

Not including certain representatives was perceived negatively by several interviewees for the risk of missing perspectives and knowledge, whereas others stated, a small panel size was preferable for the consensus process. Clinicians and methodologists deemed both, clinical and methodological expertise as beneficial, especially for specifying QI in detail to ensure content validity and feasibility. Concerns were raised by some about personal biases of panel members.

Concerning patient participation, half the interviewees stated that patient involvement was crucial. On the other hand, it was argued that most patients on the panels were not ‘vocal about measurement’ and some careful selection or prior education on measurement was necessary. Clinicians and methodologists shared similar views.

Decision-making process

There was no agreement on the ideal decision-making process. Several interviewees were advocates of formalised forms of consensus, some favouring a written Delphi

Table 2 Main topics of QI development processes with representative citations and suggested approaches as outlined by interviewees

Main topic/key message	Representative citation (capital letters (A–H) refer to the interview participant's institution, the number and small letter to clinician (1c) or methodologist (2m))	Suggested approach/area of further development
<p>(1) Organisation/context of guideline and QI development process</p> <p><i>Key message:</i> If guideline-based QI are in the scope of the guideline institution, a programmatic approach for developing QI is important. QI development and especially external implementation requires usually external cooperation and resource support depending on the institutional position</p>	<p>'And so, there's a lot of that building capacity in-house that's been happening recently; and, at the moment, it's just myself and another colleague'. A2m</p> <p>'that's quite a lot of thinking and time to set the measures up. So, I do think it needs a dedicated person who has expertise in that area, too'. G2m</p> <p>'So, if the Minister of Health thinks it's a good thing to develop quality performance indicators, there's more likely to be funding following that than if [our guideline organisation] says let's do some QPIs'. B1c</p> <p>'It depends on the financial organization because in the three guidelines which I took part in, it was different. The main guidelines and the most important of these [...] were financed by the Health Minister. So, we had money to organize everything and to involve experts of the different aspects of the disease. So, the implementation has been promoted by the Minister of Health. It was one of the examples, but in the other two it was not so because we only had finance from the (regional) ministry and very little money and so it was more <i>difficult to implement, to select, to organize and to disseminate information</i>'. E1m</p>	<p>Have a person or a team in the guideline organisation or a collaborating organisation that is responsible for the process of development of guideline-based QI</p> <p>Seek for cooperation with partners in quality improvement Adapt the QI process according to resources available for development and implementation, for example:</p> <ul style="list-style-type: none"> ▶ Consider alignment with existing QI (see 3) ▶ Limit development process to QI selection (see 5)
<p>(2) Panel composition and decision making</p> <p>a. Involvement of professionals/ education</p> <p><i>Key message:</i> Involvement of different professionals is favourable including members with methodological knowledge as well as collaboration with future implementers. The group need education and a shared understanding of the process.</p> <p>b. Patient participation</p> <p><i>Key message:</i> There are different views on patient participation of QI development and direct patient relevance of QI.</p> <p>c. Decision-making</p> <p><i>Key message:</i> There is no shared concept of a certain decision-making process. Difficulties of reaching consensus were mostly connected with feasibility aspects including expected decisions of payers</p>	<p>'So it was, we recruit a multi-disciplinary group based on who would be involved in the management of patients with [this condition]'. B1c</p> <p>'We tried to have only one representative for each professional group to avoid an over representation of one group on another'. D1m</p> <p>'I think in terms of clinical expertise; it was definitely sufficient. But in terms of measurement expertise and methodology knowledge, I think that, for us, it could improve'. A2c</p> <p>'I don't think that there were any real impediments or obstacles other than the participants actually really understanding what was being done. And that was, you know, it took a little bit of education ...'. F2c</p> <p>'I think that you couldn't do it without them [patient representatives] because the whole focus of the guideline is to be patient focused so you need to know what patients want to be able to produce relevant advice, you need to know what's important to them'. B2m</p> <p>[patients on the panels were not] '<i>vocal about measurement</i>'. A2m</p>	<p>Recruit a panel which is representative for the health professionals concerned with the respective patients, include members with methodological knowledge as well as future implementers Train the QI developing team concerning QI methodology, possibilities and limitations</p> <p>Discuss patient perspective and patient relevance of guideline-based QI in the beginning of the QI process as well as patient participation If GRADE is used in the guideline QI can be linked to prioritised patient relevant outcomes Instruct patients participating in QI process and methodology as part of the panel</p> <p><i>Area of further development/research:</i></p> <ul style="list-style-type: none"> ▶ Role and contribution of patients for guideline-based QI, <p><i>Area of further development/research:</i></p> <ul style="list-style-type: none"> ▶ Analysis of contribution of structured decision-making processes

Continued

Table 2 Continued

Main topic/key message	Representative citation (capital letters (A–H) refer to the interview participant's institution, the number and small letter to clinician (1c) or methodologist (2m))	Suggested approach/area of further development
<p>(3) QI selection criteria/attributes <i>Key message:</i> The evidence base of QI is most important and should be transparent. To assess the relevance of an QI needs knowledge of regional quality gaps/variability. Measurability as a key attribute remains a challenge. Difficulties are especially reported concerning measuring patient reported outcomes, shared decision making and individualised care for QI based on weak recommendations. To assess measurability piloting in cooperation with future implementers is favourable In terms of feasibility, already existing QI should be considered as well QI should also be accepted by clinicians</p>	<p>'A lot of the work as far as ... understanding and synthesizing the evidence has already been done during the guideline process' [...]. I think that you have to have strong evidence to make a recommendation and then on that recommendation a performance measure makes sense'. F2c</p> <p>'... one of the biggest issues that I saw with our approach to this project was coming purely from the guideline and the published peer review literature ... So, that gap in care, we still found some variability between regions, but as this was led from completely the perspective of: What does the literature say? -and not: What is the gap ... you know, I think potentially it could have been more informed and more directed if we'd been looking at what gaps exist in the first place'. F1m</p> <p>'The [guideline] has a lot of weak recommendations so we use the GRADE process, so people may want it, some people may not and it's actually a recommendation for shared decision making with physicians. So, we don't actually have a way to measure or to track shared decision-making'. G2m</p> <p>'... a lot of quality indicators have in each project to be abandoned because it's impossible with all matters actually to measure them'. D1m</p> <p>But they've given us a tremendous amount of feedback about what they need, because it's really important that we're trying not to duplicate work and develop indicators that align with what is already mandatory to report with existing systems, as well as indicators that are aligned with recommendations in the guideline that can show the implementation of the guideline. A1c</p> <p>'... it has to be achievable, as well. So there has to be something within the context of the busy, stretched, clinical services, that something is achievable both in terms of delivering it at the clinical level but also collecting data as well'. B1c</p>	<p>Use explicit evidence-based guidelines for QI development with transparent evidence base for each recommendation</p> <p>Insure to get to know regional/national quality gaps for assessing the need of an QI preferably using healthcare data, if not available, using an expert consensus</p> <p><i>Area of further development/research:</i> Integration shared decision making in quality improvement strategies and quality indicators</p> <p>Pilot QI with those who will/must implement them. Indicate lack of measurability as a rejection reason of the respective guideline-based QI recommendation in the guideline, seek for projects to explore possibilities of measurement Consider alignment with existing QI</p> <p>Consider 'resource use/expense' also for clinicians as one criterion when assessing feasibility</p>
<p>(4) Intended use and implementation <i>Key message:</i> Guideline groups (and institutions) aim their important recommendations to be implemented to ensure clinical meaningful quality improvement. If an QI is suitable for a certain purpose can only be appraised after a pilot test</p>	<p>'Trying to be ahead of the curve by identifying what [...] quality measures should be before a payer tells to do that, you know, before a payer decides what's going to be representative of our performance measure'. F2c</p> <p>'information of clinicians or politicians about recommendations translated into QI'. G2m</p> <p>'It's quite easy if you think you know beforehand how that indicator should be used, whether it should be for quality improvement, as you say, or have more formal pay-per-performance work. But I think the testing and the consultation, and the feasibility check should be what decides that, not our initial thoughts'. H2m</p>	<p>Make sure that guideline-based QI recommendations are made available to decision makers in charge of QI (see also 1)</p> <p>Do piloting to make sure the QI is suitable for the intended use</p>

GRADE, Grading of Recommendations, Assessment, Development and Evaluation; QI, quality indicator.

process, which was considered easy to understand and implement, allowing for consideration of all opinions before final consensus. However, an equal number of interviewees considered informal table discussions as sufficient, allowing for more active discussion and debate.

Clinicians and methodologists reported both difficulties in reaching consensus predominantly concerning feasibility of guideline-based QI, for example, regarding measuring shared decision making as well as acceptance

of QI due to potential economical judgements of payers or health authorities.

Collaborating with future implementers

A collaborative approach that involves feedback loops with clinical stakeholders in the field was emphasised by several interviewees as facilitating factor for QI development and implementation. Pilot testing was seen by clinicians and methodologists an essential opportunity for adapting QI based on feedback from implementers, although it was realised only in about half of the organisations due to time and resource constraints.

3. QI selection criteria/attributes

Methodologists and clinicians mentioned a wide range of QI selection criteria and attributes. Whereas interviewees of some organisations named criteria agreed by national institutions, others reported on criteria specific for their institution. Some clinicians referred to the methodologists for details.

Scientific soundness: evidence base as key

One main prerequisite for high quality QI seen by the interviewees was a strong evidence base. Clinicians and methodologists pointed out that up-to date and methodologically rigorous developed evidence-based guidelines as a basis increase QI acceptability and implementation. The GRADE (Grading of Recommendations, Assessment, Development and Evaluation) approach was deemed helpful to assess the certainty of the evidence and develop measures from recommendations with a robust evidence base. It was cautioned that gaps in current evidence as for rare diseases or weak evidence and weak recommendations themselves can hinder QI development and interpretation. In addition, some interviewees identified a time lag between the current evidence, the QI development and dissemination as a challenge which caused QI to be outdated quickly and not reflect current practice.

Relevance: need-based planning and relevance to patients

Several clinicians and methodologists considered carrying out a needs assessment prior to guideline and QI development crucial to identify key areas for improvement in care rather than focussing solely on literature driven recommendations.

QI were described mostly as structure or process related and perceived less directly relevant to patients than outcome measures. One interviewee argued that patient relevant QI would require a separate set of indicators as the physician or system perspective would differ a lot from the patient perspective. Clinicians and methodologist deemed the GRADE approach as helpful as it includes prioritising patient relevant outcomes from the beginning of guideline development. One interviewee identified a lack of resources as a hinderance to carrying out more patient oriented QI.

Measurability remains a challenge

A lack of measurability was raised by clinicians and methodologists as one of the main challenges in QI development. Over half the interviewees stated that important guideline recommendations, particularly related to qualitative and individualised aspects of care, were difficult to measure and often not reflected in current QI. Some interviewees, mostly methodologists, argued that it was difficult to develop measurable QI directly linked to patient relevant outcomes and judged the expense and effort of measuring patient reported outcomes as not feasible. A pivotal unsolved problem was seen by some clinicians and methodologists in measuring shared decision making. In line with this, several interviewees emphasised that measurable QI do not necessarily reflect the actual quality of care and should be cautiously interpreted.

In addition, a lack of infrastructure, related to IT-systems, data availability or collection capacities was reported as still hindering the development of measurable QI meeting core methodological criteria. Some interviewees also called attention to capacities of clinicians. Clinicians and methodologists argued that the use of preferably pretested existing measures can reduce the burden of data collection, save costs and guarantee that measures are feasible and valid.

Consideration of specification and pilot testing

Another challenge reported, was to develop QI that are specific but also applicable in different contexts. Clinicians and methodologists emphasised that guideline recommendations needed to be worded precisely and unambiguously to facilitate QI development. However, interviewees of one organisation decided not to determine numerator and denominator to encourage local adaptability and wide-reaching implementation whereas interviewees of other organisations argued that for an QI a clear idea of what the indicator should achieve, consideration of outliers/risk adjustment and pilot testing was needed to determine and refine the numerator and denominator.

Pilot tests were deemed by most interviewees essential for new measures or settings to improve clarity of definitions and test feasibility, validity and applicability, although only half of the interviewees stated that they were done.

4. Intended use and implementation of guideline-based QI

All interviewees named 'quality improvement' or 'quality assurance' as intended use of the QI including two mentioning 'discouragement of overuse'. Other purposes stated were information of clinicians and/or policy, identifying gaps in current care or assessment systems, assessment of guideline implementation, accreditation and benchmarking. Additionally, interviewees of half of the organisations stated reimbursement as one QI purpose. A clinician of a privately funded guideline organisation stressed the fact that a motive to develop guideline-based QI was to get quality assurance more influenced

by clinicians. He and other interviewees of public funded organisations highlighted that they could not be sure if their QI were used. One clinician did not agree with the intended use of public funded QI to centralise care by applying minimum operative volumes.

Most interviewees said that the intended use needed to be considered before indicator development. Some—clinicians and methodologists—stated, that the selection and the final use of QI should only be determined after testing, consultation and feasibility checks.

Experiences with implementation were heterogenous. Some interviewees did not see themselves responsible for implementation but deemed information of clinicians and other decision makers as their main contribution. Other interviewees were very eager to actively engage in implementation of their QI in existing systems. Those who worked in a context where QI were implemented were aware—both, methodologists and clinicians—that measurement characteristics could only be assessed after piloting.

DISCUSSION

For this qualitative study, international guideline experts have been interviewed to reveal insights into all development steps of guideline-based QI. Following a strategy of purposive sampling and using a broad definition of guideline-based QI accepting every effort to connect QI and guidelines, we could include a variety of perspectives and experiences from experts of organisations with different scope, funding and role within different healthcare systems. This allowed for the identification of common patterns, feasible approaches and factors facilitating or hindering the development of guideline-based QI. Following qualitative analysis, we identified four main topics, for which we could describe how processes were organised and experienced: (i) organisation/context of guideline and QI development process, (ii) panel composition and decision making, (iii) QI selection criteria/attributes and (iv) intended use and implementation of QI.

The G-I-N reporting standards helped to address relevant QI development aspects for the interview guide.⁸ The categorisation according to Neale¹⁹ was expedient to structure and summarise the comprehensive data.

Interviewing a clinician and a methodologist from each organisation broadened the perspective. Although some clinicians did not remember every step of the QI development process in-depth, as the guideline they had worked on was completed up to 4 years ago, our interviews revealed that clinicians and methodologists shared predominantly same views, except that clinicians had less knowledge on QI attributes and assessed measuring of direct patient relevant aspects easier than methodologists. In one case a clinician did not accept the intended use of the QI.

Our results indicate that QI development connected to guidelines can be organised in different ways depending

on the position that guideline developers have within a certain healthcare system as well as how they work together with other stakeholders. This is in line with the systematic review of Kötter *et al* who reported on different approaches published.⁷ Methodologists and clinicians from almost all organisations talked far more about organisational and context factors than we had expected. This underlines the importance of carefully setting up guideline-based QI development together with relevant stakeholder of the respective healthcare system.

Kötter *et al* found the reporting of the QI methodology often lacked transparency.⁷ The latter was still found by Becker *et al*.¹² Our results suggest that methodological weaknesses in QI development may still be due to a lack of knowledge and expertise and limited financial resources.

Our interviews with international experts unanimously revealed a need to have a dedicated person instructing guideline groups about the QI development process. This was also stressed by German guideline authors.¹¹ In contrast, Blozik *et al* had found in a survey on simultaneous development of guidelines and QI with 24 participating institutions 2012 that the initiative to develop QI was made predominantly by the guideline coordinator or group itself.²⁰ Thus, our results might reflect an organisational change towards more professionalism in terms of building institutional capacities instead of leaving the initiative to single (guideline) groups.

Interviewees deemed clinical and methodological knowledge important for an adequate QI panel. But the fact that personal biases of panel members were mentioned, highlights the need for a conflict of interest management also for QI development which should be reported on.

The varying perception interviewees had on patient involvement and patient relevance of guideline-based QI warrants further consideration. Kötter *et al* found only few papers mentioning patients as part of QI-panels.^{7 21} The recognition of patient centredness as an important dimension of quality of healthcare led to more direct patient involvement in various processes,²² but the critical voices of some interviewees reveal that there is still a need to agree on the role of patients in QI development. When asked about guideline-based QI for chronic heart failure, neither patients nor professional experts deemed the QI elected by the guideline group very patient centred although patients had been involved.^{23 24} Interviewees in our study supported the importance of 'relevance' as a selection criterion, but some highlighted that relevance can be judged differently dependent on the target group (doctors, patients or people responsible for the system). Guideline groups should reflect this and achieve to an agreement about improvement needs, avoiding a mechanical process going from guideline recommendations to QI. As stressed by some, the prioritisation of patient relevant outcomes can help solving this issue.

Most interviewees emphasised a strong evidence base as the most important criterion for guideline-based QI. Reports about flawed QIs in use for public reporting

and pay for performance²⁵ indicate a need to establish a broad and common understanding of ‘evidence’ supporting an QI. Scientific soundness of an QI should not only refer to the direct evidence base of the intervention measured (eg, antibiotic therapy for pneumonia) but also to the evidence for further QI characteristics like a required cut-off (eg, ‘door to needle time’ for antibiotics) and consideration of potential undesirable effects when implementing an QI (eg, risk of inappropriate diagnosis of pneumonia and associated risk of overtreatment with antibiotics). Scientific soundness of an QI implies additional measurement characteristics like content validity (do implementers understand the intervention in the same way) or reliability (are there systematic measurement errors). The characteristics and effects of an QI can only be confirmed by a pilot test. Interviewees—methodologists and clinicians—were mostly aware of this, but only half the organisations realised piloting.

In general, putting new measures into practice and make them measurable was—still—seen a big challenge in our study and using existing QI to avoid additional burden of documentation and validation was emphasised. This confirms the need to first appraise measures already in use when starting to develop QI from guideline recommendations.

Especially patient reported outcomes were reported as not yet feasible to measure, even though quality assurance institutions like the US National Quality Forum call since years for measuring those.²⁶

To determine the quality of the underlying evidence of a recommendation, the GRADE approach was deemed useful. Some interviewees though were uncertain how to deal with weak or conditional recommendations for which they assigned shared decision-making processes a greater weight than for strong recommendations, without satisfactory solutions to measure. Knowledge gaps in bringing together evidence-based medicine with awareness of shared decision-making processes and quality measurement were, for example, also reported in a project to implement shared decision making in the UK.²⁷ There are some survey methods described for collecting information about shared decision-making processes, such as measuring, if decision aids are distributed or using patient questionnaires,^{28 29} which need certainly further scientific development. However, the most important aspect for implementation might be the strategical decision to include shared decision making into quality improvement on the organisational or systems level with the possibility of trialability.^{29 30}

Concerning the intended use of QI, authors should agree for both, existing and new guideline-based QI, on their foreseen use and assessment should include if the QI is suitable for the intended purpose and this should be reported.⁸ QI improvement initiatives of guideline authors might be subverted, particularly when QI are used mainly to drive forward certain political aims, for example, to centralise care—a risk that may be dependent on the funding source of the respective organisation.

This study has some limitations: first, transcripts or results were not returned to participants for comment or correction. However, we had very clear transcriptions after reviewing doubtful passages with the recorded interviews. Another limitation is, that despite looking for maximum variation, most organisations identified and willing to participate were public funded and acted on the national level with location in Europe or North America. Due to resource constraints the number of interviews was limited, thus it cannot be ensured that we captured all important aspects for each theme, especially concerning the processes and experiences of private funded and local/regional organisations. Nevertheless, we have found an instructive sample concerning the range of views on guideline-based QI.

CONCLUSION

Our study indicates that a diversity of development methods, realisation and use of guideline-based QI exists and development of guideline-based QI can succeed either parallel to or following guideline development. The results suggest that success will need:

- ▶ A programmatic/strategic decision to develop and implement guideline-based QI with a dedicated person to lead the process.
- ▶ The establishment of an adequate panel including clinicians and methodologists to develop guideline-based QI.
- ▶ The provision of methodological instruction on the development of guideline-based QI.
- ▶ The need to consider methodological criteria for QI, such as relevance, evidence base and feasibility.
- ▶ Collaboration and the establishment of strategic partnerships to implement guideline-based QI in a health-care system.

Further exploration and development are needed concerning the role of patients, direct patient relevance of QI, conflict of interest management of QI panels as well as measurement of qualitative aspects as shared decision making.

Author affiliations

¹Institute for Medical Knowledge Management c/o Philipps University Marburg, Association of the Scientific Medical Societies in Germany, Marburg/Berlin, Germany

²Fakultät Gesundheits- und Pflegewissenschaften, Westsächsische Hochschule Zwickau, Zwickau, Germany

³Center for Evidence-based Healthcare, Faculty of Medicine Carl Gustav Carus, Technical University, Dresden, Germany

⁴Institut für Forschung in der Operativen Medizin, University Witten Herdecke, Witten, Germany

Acknowledgements We would like to thank all participants of the interviews for their time and valuable contributions. We would further like to thank Professor Dr Max Geraedts for methodological advice and supervision of the whole project and him and Dr rer. -medic. Susanne Blödt for the review of this publication.

Contributors The study was mainly carried out by three researchers MN, MB and MS (MN, MS, MB: study planning, study protocol, interview guide; MB: leading interviews; MN, MB: analysis, manuscript preparation) see also the Researcher characteristics section. The other authors were involved in study planning (JS,

SD), reviewed the study protocol (KA, MBe, SD, IK, JS), the interview guide and the manuscript (both all authors: MN, MB, MS, KA, MBe, SD, IK, JS).

Funding The study was funded by the German Research Foundation (DFG No. 289625106).

Competing interests MN and IK are employed by the AWMF as methodological advisers for guideline and guideline-based QI development. The AWMF receives a continuous grant from the German Cancer Aid for guideline and guideline-based QI development.

Patient consent for publication Not required.

Ethics approval Ethical approval for the study was given by the ethics commission of the Medical Chamber Berlin/Germany (Berlin Chamber of Physicians No. 31-17).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The datasets used and analysed during the current study are available from the corresponding author on request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Monika Nothacker <http://orcid.org/0000-0003-4928-7391>

REFERENCES

- Baker DW, Qaseem A, Reynolds PP, *et al.* Design and use of performance measures to decrease low-value services and achieve cost-conscious care. *Ann Intern Med* 2013;158:55–9.
- Nothacker M, Muehe-Borowski C, Kopp IB. [Measuring quality in the German Guideline Programme in Oncology (GGPO)—methodology and implementation]. *Z Evid Fortbild Qual Gesundheitswes* 2014;108:470–80.
- Follmann M, Schadendorf D, Kochs C, *et al.* Quality assurance for care of melanoma patients based on guideline-derived quality indicators and certification. *J Dtsch Dermatol Ges* 2014;12:139–47.
- Lawrence M, Olesen F. Indicators of quality in health care. *Eur J Gen Pract* 1997;3:103–8.
- Schmitt J, Petzold T, Deckert S, *et al.* [Recommendations for quality indicators in German S3 guidelines: a critical appraisal]. *Gesundheitswesen* 2014;76:819–26.
- Liang L, Abi Safi J, Gagliardi AR, *et al.* Number and type of guideline implementation tools varies by guideline, clinical condition, country of origin, and type of developer organization: content analysis of guidelines. *Implement Sci* 2017;12:136.
- Kötter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators—a systematic review. *Implement Sci* 2012;7:21.
- Nothacker M, Stokes T, Shaw B, *et al.* Reporting standards for guideline-based performance measures. *Implement Sci* 2016;11:6.
- Schmitt J, Deckert S, Arnold K. *Evidenz- und konsensbasierter standard für die Entwicklung von leitlinienbasierten Qualitätsindikatoren*, 17 Deutscher Kongress für Versorgungsforschung (DKVF). German Medical Science Publishing House, 2018.
- Deckert S, Steudtner M, Becker M. (Wie) erfolgt die Ableitung von Qualitätsindikatoren zur Messung und Bewertung der Versorgungsqualität im Rahmen von S3-Leitlinien? Eine Übersichtsarbeit [(How) are quality indicators for measuring and appraising the quality of healthcare derived from evidence-based clinical practice guidelines? A review]. *Z Evid Fortbild Qual Gesundheitswes* 2019.
- Arnold K, Breuing J, Becker M, *et al.* [Development of guideline-based quality indicators: a qualitative study on barriers and facilitating factors from the perspective of S3-guideline authors]. *Z Evid Fortbild Qual Gesundheitswes* 2019;147-148:34–44.
- Becker M, Breuing J, Nothacker M, *et al.* Guideline-based quality indicators—a systematic comparison of German and international clinical practice guidelines. *Implement Sci* 2019;14:71.
- Mayring P. *Qualitative Inhaltsanalyse, Grundlagen und Techniken, 12. Auflage*, Beltz Verlag, Weinheim und Basel, 2015.
- Stedtner M, Kopp I, Arnold K. *Experiences with methodological requirements for guideline-based performance measures and their practical application. Evidence map and protocol for a qualitative study. Poster, Global Evidence Summit. Kapstadt 2017, 2017.*
- Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–57.
- Patton MQ. *Qualitative research*. Chichester: John Wiley & Sons, Ltd, 2005.
- Schreier M. 3.2.2. Qualitative Stichprobenpläne. In: Mey G, Muck K, eds. *Handbuch qualitative Forschung in der Psychologie*, VS Verlag, Wiesbaden, 2010: 245–7.
- MAXQDA. Standard version for windows, 2017. Available: <https://www.maxqda.com/> [Accessed 18 Dec 2019].
- Neale J. Iterative categorization (IC): a systematic technique for analysing qualitative data. *Addiction* 2016;111:1096–106.
- Blozik E, Nothacker M, Bunk T, *et al.* Simultaneous development of guidelines and quality indicators -- how do guideline groups act? A worldwide survey. *Int J Health Care Qual Assur* 2012;25:712–29.
- Kötter T, Schaefer FA, Scherer M, *et al.* Involving patients in quality indicator development - a systematic review. *Patient Prefer Adherence* 2013;7:259–68.
- Brett J, Staniszewska S, Mockford C, *et al.* Mapping the impact of patient and public involvement on health and social care research: a systematic review. *Health Expect* 2014;17:637–50.
- Baudendistel I, Noest S, Peters-Klimm F, *et al.* Bridging the gap between patient needs and quality indicators: a qualitative study with chronic heart failure patients. *Patient Prefer Adherence* 2015;9:1397–405.
- Pohontsch NJ, Herzberg H, Joos S, *et al.* The professional perspective on patient involvement in the development of quality indicators: a qualitative analysis using the example of chronic heart failure in the German health care setting. *Patient Prefer Adherence* 2015;9:151–9.
- Wachter RM, Flanders SA, Fee C, *et al.* Public reporting of antibiotic timing in patients with pneumonia: lessons from a flawed performance measure. *Ann Intern Med* 2008;149:29–32.
- National Quality Forum. *Patient reported outcomes (PROs) in performance measurement*. Washington DC, 2013.
- Joseph-Williams N, Lloyd A, Edwards A, *et al.* Implementing shared decision making in the NHS: lessons from the magic programme. *BMJ*;2017;j1744.
- Barr PJ, Thompson R, Walsh T, *et al.* The psychometric properties of collaborate: a fast and frugal patient-reported measure of the shared decision-making process. *J Med Internet Res* 2014;16:e2.
- Scholl I, LaRussa A, Hahlweg P, *et al.* Organizational- and system-level characteristics that influence implementation of shared decision-making and strategies to address them - a scoping review. *Implement Sci* 2018;13:40.
- Forcino RC, Meinders MJ, Engel JA, *et al.* Routine patient-reported experience measurement of shared decision-making in the USA: a qualitative study of the current state according to frontrunners. *BMJ Open* 2020;10:e037087.