



Evidence-basing for quality improvement; bringing clinical practice guidelines closer to their promise of improving care practices

Marjolein Moleman¹ | Sonja Jerak-Zuiderent² | Hester van de Bovenkamp³ | Roland Bal³ | Teun Zuiderent-Jerak¹

¹Faculty of Science, Athena Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

²Department of Ethics, Law and Humanities, Amsterdam University Medical Centre (Location AMC), Amsterdam, The Netherlands

³Institute of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

Correspondence

Teun Zuiderent-Jerak, Faculty of Science, Athena Institute, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands.
Email: teun.zuiderent-jerak@vu.nl

Funding information

Dutch Council for Quality of Healthcare (Regieraad), The Hague, The Netherlands

Abstract

Background: Clinical practice guidelines (CPGs) have become central to efforts to change clinical practice and improve the quality of health care. Despite growing attention for rigorous development methodologies, it remains unclear what contribution CPGs make to quality improvement.

Aim: This mixed methods study examines guideline quality in relation to the availability of certain types of evidence and reflects on the implications of CPGs' promise to improve the quality of care practices.

Methods: The quality of 62 CPGs was assessed with the Appraisal of Guidelines, Research, and Evaluation (AGREE) instrument. Findings were discussed in 19 follow-up interviews to examine how different quality aspects were considered during development.

Results: The AGREE assessment showed that while some quality criteria were met, CPGs have limited coverage of domains such as stakeholder involvement and applicability, which generally lack a 'strong' evidence base (e.g., randomized controlled trials [RCT]). Qualitative findings uncovered barriers that impede the consolidation of evidence-based guideline development and quality improvement including guideline scoping based on the patient-intervention-comparison-outcome (PICO) question format and a lack of clinical experts involved in evidence appraisal. Developers used workarounds to include quality considerations that lack a strong base of RCT evidence, which often ended up in separate documents or appendices.

Conclusion: Findings suggest that CPGs mostly fail to integrate different epistemologies needed to inform the quality improvement of clinical practice. To bring CPGs closer to their promise, guideline scoping should maintain a focus on the most pertinent quality issues that point developers toward the most fitting knowledge for the question at hand, stretching beyond the PICO format. To address questions that lack a strong evidence base, developers actually need to appeal to other sources of knowledge, such

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Evaluation in Clinical Practice* published by John Wiley & Sons Ltd.

as quality improvement, expert opinion, and best practices. Further research is needed to develop methods for the robust inclusion of other types of knowledge.

KEYWORDS

clinical guidelines, epistemology, evidence-based medicine, healthcare

1 | INTRODUCTION

Since the rise of evidence-based medicine (EBM), clinical practice guidelines (CPGs) have become the tool of choice for narrowing the gap between external evidence and the practice of care delivery. CPGs have therefore become central to efforts to change clinical practice and improve the quality of health care.¹ Parallel to the burgeoning industry of evidence-based guideline development, however, there have been growing concerns about the extent to which CPGs can live up to the promise of serving as instruments for quality improvement.¹ Much work has been done to formalize the guideline development process,²⁻⁴ which has gained priority to address significant variations in guideline recommendations and quality.⁵⁻⁷ Standardizing the quality of reporting aims to improve the process of searching, selecting, and presenting evidence, which has an indirect relationship to improving the quality of the guidance itself.

Despite the introduction of manuals to guide and evaluate the quality of the guideline development process, studies show low performance on several quality domains such as stakeholders involvement, cost implications, and applicability to clinical practice,^{5,6,8,9} domains which generally lack a 'strong' evidence base (e.g., randomized controlled trials [RCT]). Knaapen¹⁰ points to the need to examine how the absence of evidence is managed in evidence-based guideline development practices and the implications for CPGs as tools for quality improvement. Previous studies suggest a discrepancy between the guideline development process and final recommendations.^{11,12} An ethnographic study of guideline development practices found that guideline panel members do not restrict themselves to searching, appraising, and presenting research evidence, but entwine this evidence with several repertoires, referred to as the repertoires of process, politics, and practice.¹¹ These repertoires allow guideline developers to consider future practices and users during development, hence helping them translate research evidence to specific care practices in context. In spite of the use of such repertoires, the end-product—the CPG—remained very technical.¹¹

Authors^{6,13} suggest that the suboptimal quality of CPGs could negatively affect utilization and adherence to CPGs. Even if the evidence base is solid or the methods used to create the CPG are of high quality, clinical practice guidelines may still yield recommendations that are difficult to use or implement.¹⁴ Beyond significant methodological improvements in guideline development improvements over the past decades, addressing quality concerns especially increasing the applicability of CPGs to everyday clinical practice is still needed to adequately address quality improvement needs.^{6,13,15}

This study that examined why evidence-based CPGs remain removed from their promise to serve as instruments for quality

improvement was conducted in 2010. The threefold aim of this paper is to examine (1) how a large set of Dutch guidelines score on the AGREE domains, (2) how the content of these CPGs is shaped by the availability of certain types of evidence, and (3) what this means for the role of CPGs as tools for quality improvement.

Ongoing discussions regarding the narrow definition of evidence in EBM, which tends to neglect important sources of knowledge thereby impairing quality improvement of care practices,¹⁶⁻¹⁹ underline the topical relevance of this study. To our knowledge, this study remains the first and only study to empirically examine the influence of dominant systems of guideline development on guideline quality and contributes to the discourse regarding EBM and quality improvement by formulating recommendations on how to bring evidence-based CPGs and their development infrastructures closer to their promise of improving health care quality.

1.1 | Dominant systems of guideline development

The field of guideline development is highly dynamic. In a review of trends in guideline development practices, Kredo et al.⁴ describe that 'transparently constructed evidence-informed approaches integrated with expert opinion and patient values have rapidly gained acceptance over the past two decades as the best approach to CPG development'.^{4,p123} To standardize development practice, a range of tools and checklists have been developed, such as institutional standards (e.g., SIGN, NICE, WHO Handbook for Guideline Development), (customized) standards used by speciality organizations (e.g., AAN Development Process), and checklists offered by international guideline networks (e.g., GIN-McMaster Checklist, RAND UCLA Appropriateness Methodology). Furthermore, tools to evaluate the quality of CPGs (e.g., the Appraisal of Guidelines Research and Evaluation [AGREE] instrument) may also be used by guideline developers to structure the development process.

Depictions of evidence-based guideline development found in literature commonly distinguish three phases:

- (1) *Preparation*: setting up a guideline panel that defines the PICO²⁰ and formulates key questions. Attention points at this stage include: assembling a multidisciplinary guideline panel to ensure representation of relevant stakeholders and the mitigation of conflicts of interest within the panel.
- (2) *Evidence synthesis*: searching and appraising (systematic reviews of the) evidence using tools such as AMSTAR 2²¹ and the ROBIS tool.²² Attention points are the transparent and rigorous application of

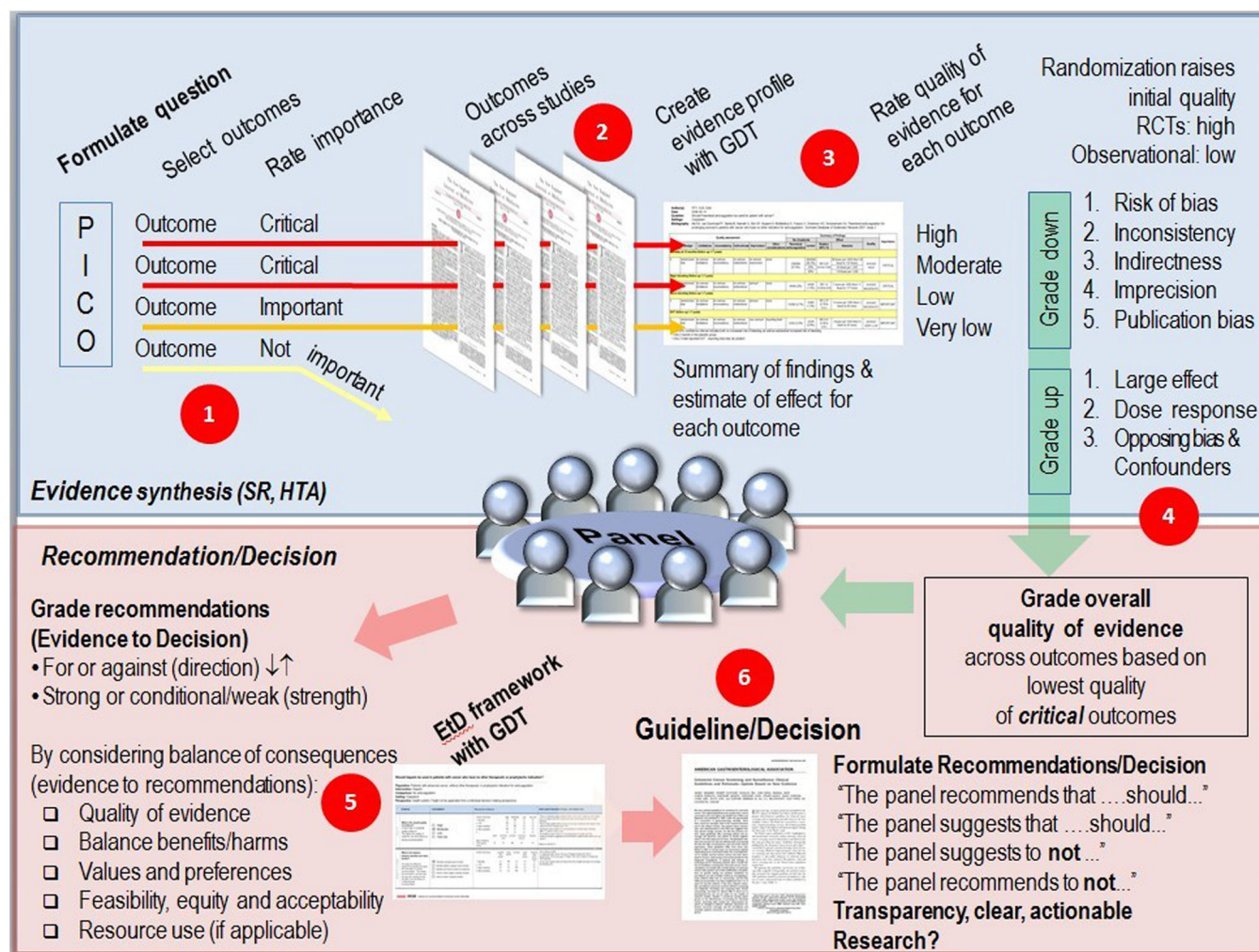


FIGURE 1 Depiction of the phases of guideline development: the phase of preparation (Phase 1), evidence synthesis (Phase 2–4) including the assessment of the uncertainty of the evidence following the GRADE criteria for (Phase 3) and the formulation of recommendations (Phase 6) including the GRADE Etd criteria (Phase 5) for integrating the best available evidence to answer the question at hand²³

evidence procedures, often guided by tools to assess the certainty of summarized evidence such as Grading of Recommendations Assessment, Development, And Evaluation (GRADE).²³ After the outcomes across studies have been compiled, the quality or certainty of the evidence, "the extent to which our confidence in an estimate of the effect is adequate to support a particular recommendation"²⁴ is assessed. Evidence from RCTs starts at high confidence and, because of residual confounding, evidence from observational data starts at low confidence. Confidence in estimates may be upgraded or downgraded based on a set of criteria.²⁵ GRADE is used as an example (Figure 1) as it is the most used approach to assess the certainty of the evidence, still many organizations do not use GRADE because it may not be applicable or feasible. Finally, conclusions are drawn that summarize the evidence and the quality of that evidence.

(3) *Formulating recommendations*: to arrive at recommendations for practice, panels integrate the best available evidence to answer the question at hand. GRADE evidence-to-decision (Etd)²⁷ is an example of a framework to structure the process of using evidence to inform health care decisions. The Etd framework outlines different criteria (depicted in Figure 1, Phase 5) to make judgements about the pros and cons of interventions being considered; ensuring that panel members consider all the important factors for making a decision, and providing panel members with a concise summary of the best available evidence about each criterion to inform their judgements.²⁸ In a series of papers on 'the evolution of GRADE', Mercuri and Gafni²⁹ highlight considerable ambiguity regarding the operationalization and integration of the GRADE Etd criteria when developing recommendations. Developers of the Etd framework indeed note that panels must consider the implication and importance of each judgement.³⁰ Knowledge integration in this final phase thus remains a complex, social process since guideline panels need to

*Depicted in Phase 4 of Figure 1.

consider, discuss and judge how different sources of (possibly conflicting) knowledge can be integrated to inform final recommendations.³¹

These three phases of evidence-based guideline development formed the backdrop for our study.

2 | METHODS

We conducted a mixed-methods study consisting of two phases: a quantitative evaluation of CPG documents and qualitative follow-up interviews.

2.1 | Quantitative analysis of CPGs

First, we examined the quality of 62 CPGs, to explore how the content of these CPGs is shaped by the availability of certain types of evidence. We included CPGs from the 25 most prevalent conditions in the Netherlands, drawn from the 'top 100 CPGs database'[†] of the Dutch Council for Quality of Health care. CPGs were excluded if they diverted from the 'traditional' clinical practice guideline format, such as 'primary care collaboration agreements'. In case a monodisciplinary CPG had been updated to a multidisciplinary guideline, only the latter was included for evaluation. Finally, international guidelines were excluded because the project commissioner planned on using study findings to develop a manual to streamline development practices in the Netherlands.

2.2 | AGREE instrument

The quality of 62 CPGs was evaluated using an extended version of the Appraisal of Guidelines Research and Evaluation (AGREE) instrument.^{32,33} The AGREE instrument assesses the process of guideline development and reporting of this process. AGREE consists of 23 items (Table 1, left-side column) comprising six quality-related domains: scope and purpose, stakeholder involvement, rigour of development, clarity of presentation, applicability, and editorial independence. Given our focus on bringing CPGs closer to their promise of serving as quality instruments, additional items were included (Table 1, right-side column). The AGREE domain *stakeholder involvement* was extended with items on *patient involvement* and *user involvement* to distinguish between stakeholders that may have different priorities regarding quality improvement. The *applicability* domain was extended with items on the *practice of health care organization* and *cost implications*. Finally, items were included to assess reporting of *safety risks*. The additional items were the result of several brainstorm sessions between

the research team and guideline development experts from the commissioning organization. To validate the extended AGREE instrument, it was reviewed and tested by two guideline developments experts and one health care quality expert. Based on their comments and suggestions the evaluation instrument was adjusted and refined.

The AGREE manual recommends that the evaluation of CPGs be conducted independently by four evaluators based on which a mean AGREE score is calculated. Four researchers conducted the AGREE evaluation. The researchers had no prior experience using AGREE to evaluate CPGs. To increase the reliability and validity of the evaluation, the researchers were selected based on their expertise: two experts in health care economics and health technology assessment evaluated the AGREE domains Methodology, Clarity of Presentation and items about cost-effectiveness within the Applicability domain. Experts were selected based on their specific topic expertise. Two experts in health care governance evaluated the AGREE domains Scope and Purpose, User Involvement, Editorial Independence and additional AGREE items (Table 1, right-side column). By involving researchers with topic expertise and assigning AGREE domains that fall within their field of expertise, the quality of the evaluation was strengthened. To safeguard the reliability of evaluations, the evaluators discussed their scoring method and checked for large differences (>1) to see whether a consensus agreement needed to be reached. The scores were documented in Microsoft Excel following the AGREE procedures.

2.3 | Follow-up interviews

After the evaluation of the 62 CPGs was completed, follow-up interviews were conducted, to examine which considerations informed the inclusion and exclusion of certain quality aspects in CPGs and what this means for the role of CPGs as tools for quality improvement. Three trained research interviewers conducted individual semi-structured interviews of approximately 45–60 min. Two topic guides (supplement) were used to discuss how the guideline had been developed, how different quality aspects were dealt with, how quantitative findings could be explained, and how guidelines were received by end-users.

A total of 13 developers were interviewed to discuss their experiences with the development process. The guideline developers were recruited based on their involvement in one of six recently developed CPGs that were selected for further analysis in close collaboration with the project commissioner. Furthermore, the selection included a mixture of guideline scope (monodisciplinary/multidisciplinary) and medical disciplines. In addition, a purposive sample of six experts was invited via email to participate based on their longstanding national and international involvement in the field of guideline development. Interviews with experts discussed their perspectives on the dominant systems and tools in the field of guideline development, their perspectives on the incorporation of guidelines' applicability to clinical practice during the development process, and their perspectives on the role of different

[†]The top 100 CPG database has been replaced by a repository containing all Dutch clinical practice guidelines that is available at <https://richtlijnendatabase.nl/>.

**TABLE 1** AGREE items and additional items used to evaluate the quality of 62 CPGs

Domains	AGREE I items	Additional items
Scope and purpose	<ol style="list-style-type: none"> 1. The overall objective(s) of the guideline is (are) specifically described. 2. The clinical question(s) covered by the guideline is (are) specifically described. 3. The patients to whom the guideline is meant to apply are specifically described. 	
Stakeholder involvement	<ol style="list-style-type: none"> 4. The guideline development group includes individuals from all the relevant professional groups. 5. The patients' views and preferences have been sought. 6. The target users of the guideline are clearly defined. 7. The guideline has been piloted among target users. 	<p><i>Patient involvement</i></p> <ol style="list-style-type: none"> 5a. Patient participation was part of the guideline development process. 5b. Patient's input is clearly described. 5c. Patients' access to the guideline is addressed. <p><i>User involvement</i></p> <ol style="list-style-type: none"> 7a. Pilot results are included in the guideline. 7b. Guideline's access to different types of users is described.
Rigour of development	<ol style="list-style-type: none"> 8. Systematic methods were used to search for evidence. 9. The criteria for selecting the evidence are clearly described. 10. The methods used for formulating the recommendations are clearly described. 11. The health benefits, side effects and risks have been considered in formulating the recommendations. 12. There is an explicit link between the recommendations and the supporting evidence. 13. The guideline has been externally reviewed by experts before its publication. 14. A procedure for updating the guideline is provided. 	<p><i>Safety</i></p> <ol style="list-style-type: none"> 11a. Safety risks and adverse events were explicitly searched for. 11b. Identified safety risks and adverse events were considered in the formulation of recommendations.
Clarity of presentation	<ol style="list-style-type: none"> 15. The recommendations are specific and unambiguous. 16. The different options for management of the condition are clearly presented. 17. Key recommendations are easily identifiable. 18. The guideline is supported with tools for application. 	
Applicability	<ol style="list-style-type: none"> 19. The potential organizational barriers in applying the recommendations have been discussed. 20. The potential cost implications of applying the recommendations have been considered. 21. The guideline presents key review criteria for monitoring and/or audit purposes. 	<p><i>Organizing and coordinating healthcare services</i></p> <ol style="list-style-type: none"> 19a. Collaboration agreements are explicitly addressed in the guideline. 19b. Opportunities for care substitution are explicitly addressed in the guideline. <p><i>Cost implications</i></p> <ol style="list-style-type: none"> 20a. The cost-effectiveness questions are clearly formulated. 20b. Outcome measures are clearly identified, measured and evaluated. 20c. All relevant costs of alternatives are measured. 20d. There is explicit attention for the financial implications of medical technology use (costs and benefits).
Editorial independence	<ol style="list-style-type: none"> 22. The guideline is editorially independent from the funding body. 23. Conflicts of interest of guideline development members have been recorded. 	

TABLE 2 Interview participants

(Past) Occupation	Role in guideline development	Involved in CPG development (n = 62)?
Guideline experts (n = 6)		
General practitioner	Coordinator	Yes
General practitioner	Researcher, guideline developer	No
Professor	Chairperson and guideline developer	No
Professor, MD	Guideline developer	No
Physician	Department chairperson Cochrane Collaboration	No
Policymaker	Medical policy coordinator (government)	No
Guideline developers (n = 13)		
Professor, general practitioner	Chairperson	Yes
General practitioner, researcher	Secretary	Yes
Physical therapist, researcher	Secretary	Yes
Physician	Chairperson, editor	Yes
Researcher	Project leader	Yes
Psychiatrist	Chairperson	Yes
Researcher	Author	Yes
Nurse	Panel member	Yes
General practitioner	Panel member	Yes
Professor, MD	Chairperson	Yes
Psychiatric nurse	Chairperson, editor	Yes
Policy officer guideline development	Project leader	Yes
Director	Coordinator	Yes

quality aspects (e.g., stakeholder involvement, resource implications, and safety) in guideline development. Consent for the interviews was confirmed upon agreement to participate in this study. Interviewees were informed about the study objective at the start of the interviews and were asked to agree to the proposed terms of the interview verbally. The anonymity of all study participants was guaranteed.

Interview participants had a background in general medical practice (n = 5), oncology (n = 2), psychiatry (n = 5), neurology (n = 2) and social medicine (n = 1). Table 2 provides an overview of participant characteristics.

With participant consent, interviews were audio-recorded and transcribed verbatim. All the interviews were summarized and sent to the respondent for member check. Respondents were asked to supply changes and additions via email. Summaries and subsequent analyses were adjusted accordingly. Recruitment, data collection, qualitative analysis were concurrent activities. The interviewers met regularly to develop an iterative and shared understanding of the data and to identify attention points for follow-up interviews. To triangulate findings, interviewees were asked to reflect on responses given by other interview participants, when similar topics came up during the interview.

Each interview transcript was analyzed by the principal interviewer using Atlas.ti software. Coding followed a content analysis approach,³⁴ in which preidentified codes based on the domains of the AGREE instrument and additional quality topics (patient involvement, user involvement, the practice of health care organization, cost implications, and safety risks) were used. Furthermore, inductive coding for newly identified themes was used. Finally, reflection sessions with the project commissioner were organized to collect feedback from experts in the field of guideline development and to validate the findings.

3 | RESULTS

3.1 | Quantitative analysis of 62 CPGs

Overall, the CPGs scored high on *Scope and Purpose*, *Clarity of Presentation*, and *Rigour of Development*, as shown in Figures 2–4. The scores indicate that rigorous methods for searching and reviewing the scientific evidence and grading the strength of the evidence have been fairly well established within the guideline development process. Scores on the domain of *Editorial independence* (Figure 5)



Scope and purpose

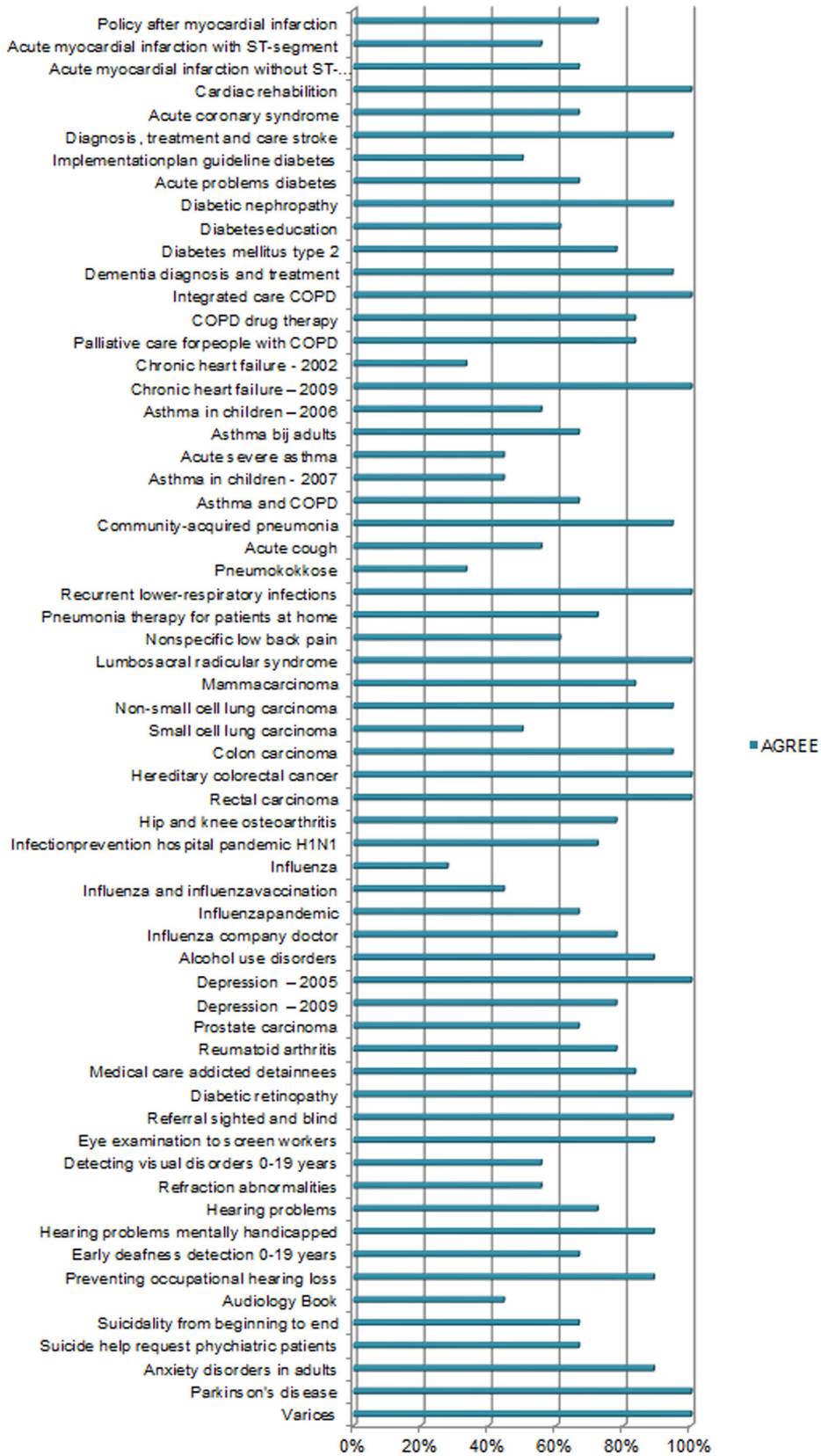


FIGURE 2 AGREE scores on Scope and Purpose. The figure depicts the total score of the items relating to one AGREE domain. A score of 4 [strongly agree] on all items would lead to a 100% score in the figure



Clarity and presentation

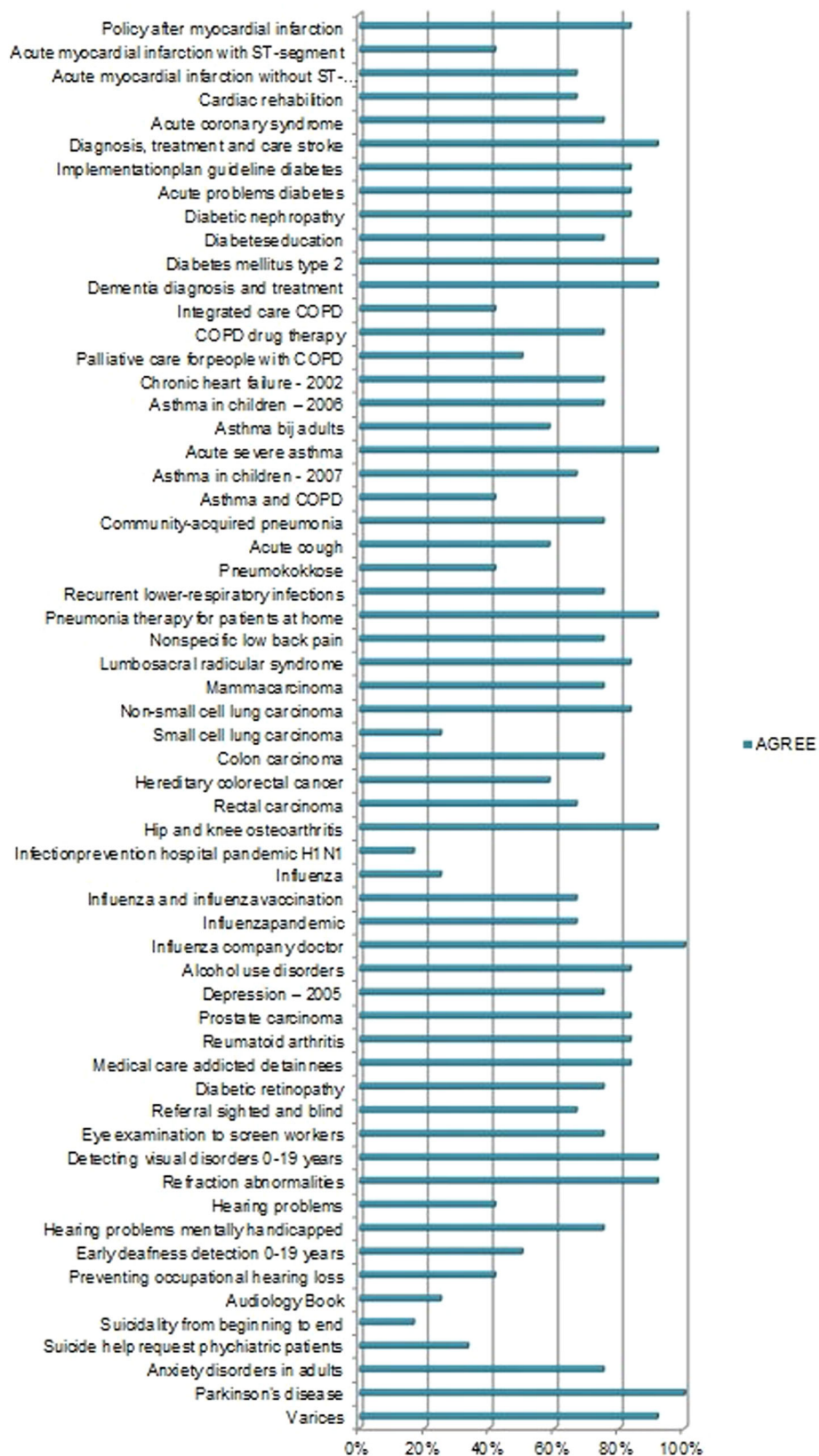


FIGURE 3 AGREE scores on Clarity of Presentation. Two guidelines were missing in the data file due to a data error. Figure 3 reports on the Clarity of Presentation of 60 CPGs



Rigour of development

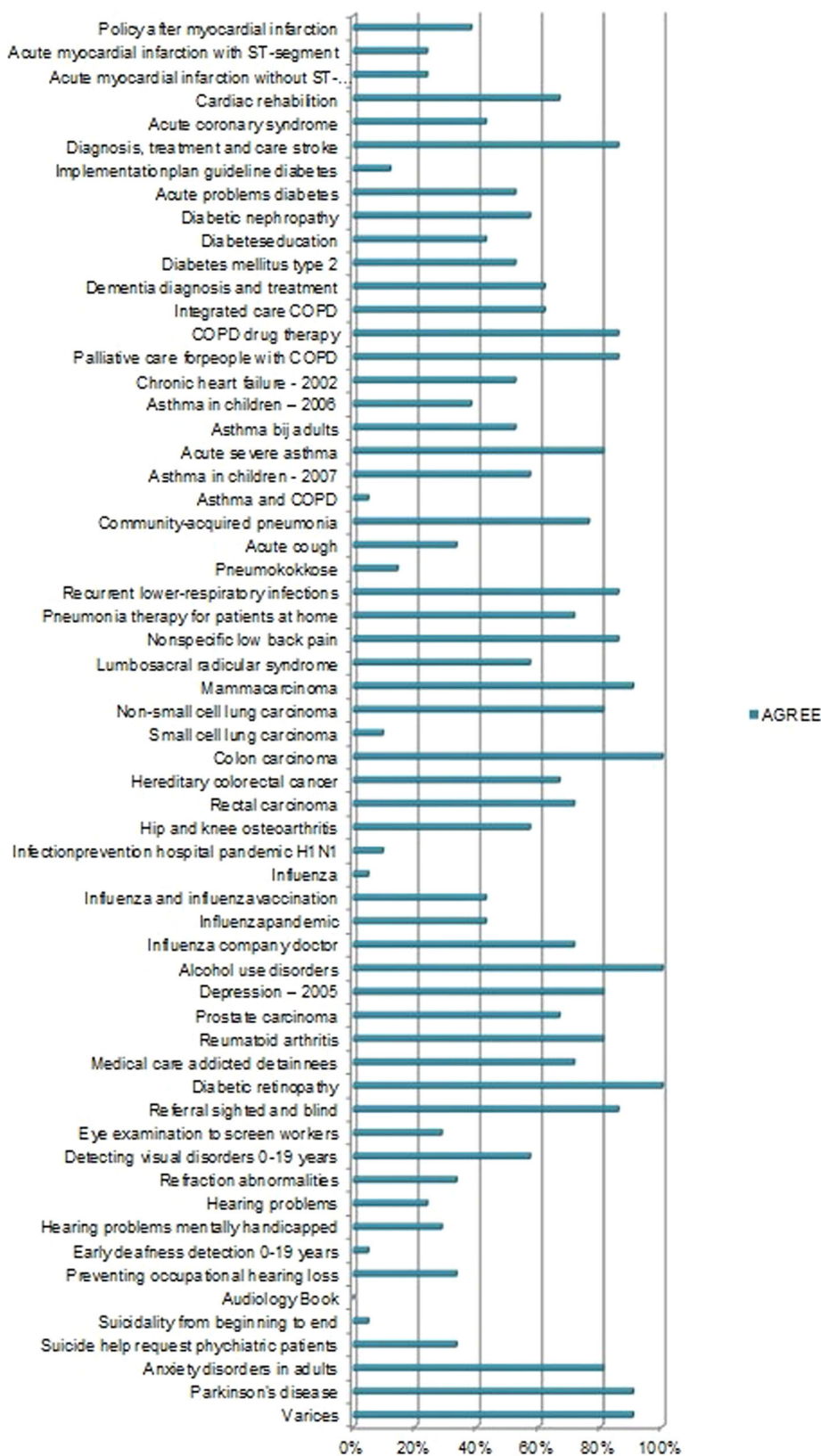


FIGURE 4 AGREE scores on Rigour of Development



Editorial independence

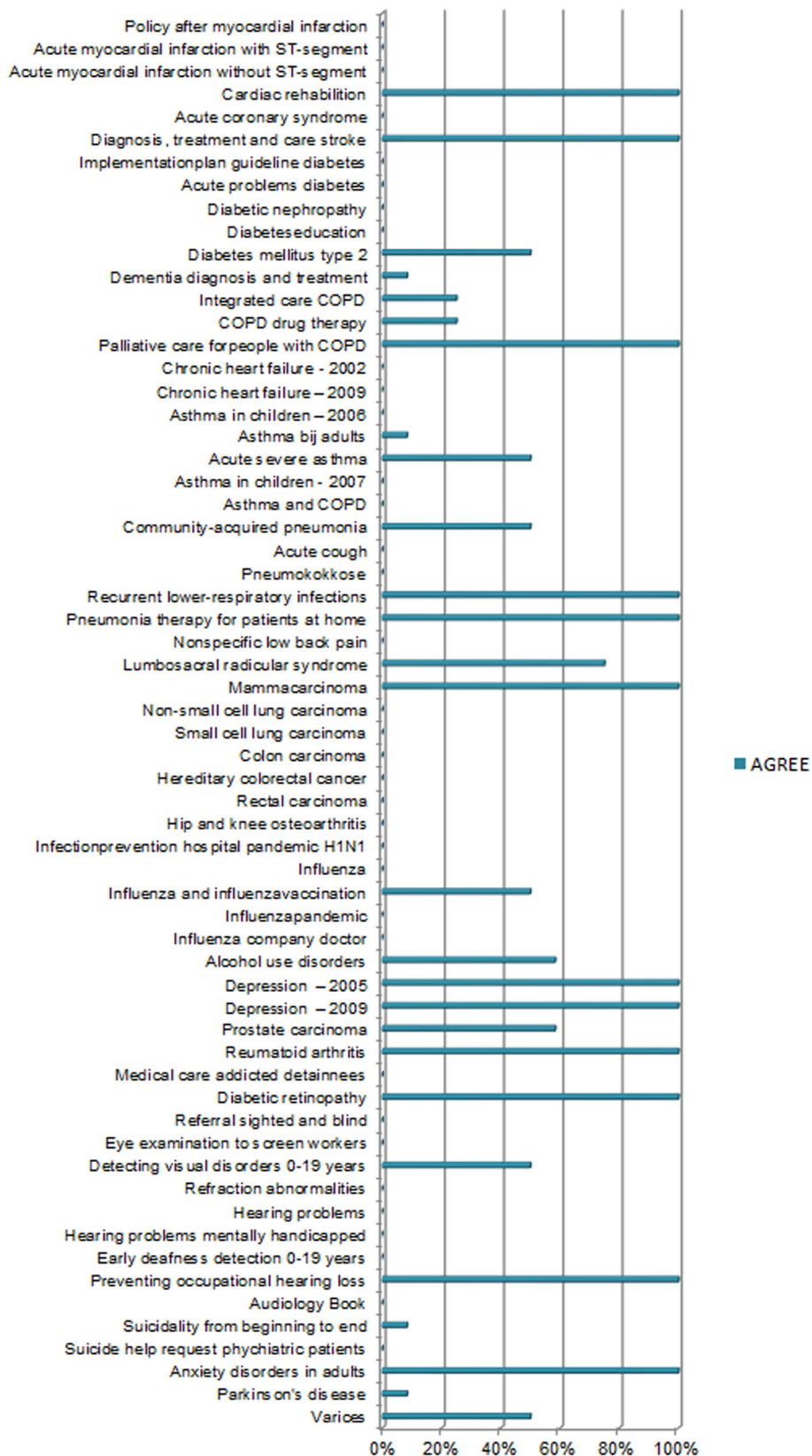


FIGURE 5 AGREE scores on Editorial independence. Scores on the domain of Editorial independence show an 'all or nothing' pattern across CPGs. Other studies^{5,6,8,9,35} similarly show relatively lower scores on this domain. 0% scores on this domain do not necessarily indicate a lack of editorial independence. It shows that the procedural criteria needed to explicate 'editorial independence' were not included in the guideline document



showed an 'all or nothing' pattern across CPGs. 0% scores on this domain do not necessarily indicate a lack of independence. It shows that the procedural criteria needed to explicate editorial independence were not included in the guideline document. Figures 6 and 7 illustrate that CPGs generally scored lower on domains of *Stakeholder involvement* and *Applicability*, which may suggest that rigorous methods for evidence appraisal take precedence over practical considerations of implementation and use.

3.2 | Additional items

3.2.1 | Patient involvement

The AGREE domain *Stakeholder involvement* was extended with three items on patient involvement (Table 1). Figure 8 shows overall low scores on this topic. Even if the patient's views were sought, its impact on the content of the guideline, for example, adjusting the formulation of a guideline recommendation, was rarely explicated in the CPG. This is not to say it did not have an impact. It may also suggest that it is difficult to explicate experiential patient knowledge in the CPG. Findings also show that CPGs scoring low on patient involvement may still be endorsed by affiliated patient organizations as sufficiently patient-centred.

3.2.2 | User involvement

Additional items regarding *user involvement* examined whether targets users and feedback from health care professionals were incorporated in the guideline. Scores on Item 7 show that only 10% of CPGs reported on the conduct of pilot tests. Figure 9 shows that feedback from the field is rarely explicated in the guideline. An interview respondent described that low use of pilot tests in guideline development motivated the removal of this item from the stakeholder involvement domain when AGREE I was updated to AGREE II. Although pilot tests are still mentioned in AGREE II as an option to assess the applicability of the guidance, the conduct of a pilot test is no longer part of the AGREE scoring system.

3.2.3 | The practice of health care organization

The AGREE domain *Applicability* was extended with two items[‡] on (1) collaboration agreements and 2) opportunities for care substitution. Figure 10 shows that the CPGs score quite differently on this topic: whereas some CPGs pay significant attention to this topic, the majority of CPGs mention it, without much detail.[§] Furthermore, findings show that few CPGs explicate possibilities to substitute care, with the

exception of the guideline Parkinson's disease that emphasizes the role of specialized nurses.

3.2.4 | Cost implications

Only one-third of CPGs explicitly reported taking cost implications into account during development. Cost implications may be excluded in the absence of cost-effectiveness studies. Alternatively, costs may be considered in terms of the (financial) feasibility of recommendations (e.g., AGREE item 19). Figure 11 shows that, apart from one outlier that scored high on this domain (75%), other CPGs scored low on this domain.

3.2.5 | Safety

Figure 12 suggests that the consideration of safety plays an important role in guideline development, with 66% of the CPGs scoring above 60%. Although safety has long been an integral part of clinical work, the focus on safety in clinical research and the scholarly discourse has catalyzed since the start of the new millennium.

3.2.6 | Qualitative analysis

This section reports on qualitative findings regarding the guideline development process. The findings are structured according to the phases depicted in dominant systems of guideline development (Figure 1). Respondents tried to bring considerations of the clinical encounter into guideline development, for example by blurring the lines between what is considered 'the work of researchers (methodologists)' and what is considered 'the work of panel members' and creating workarounds to avoid tensions between them. This section compares the formal system of guideline development and the 'informal processes' that illustrate how panel members negotiate procedural rules, methodologically robust procedures, and different considerations of quality to arrive at recommendations for clinical practice.

3.2.7 | Preparation phase

Since CPGs are intended for use in health care decision-making, guideline development groups need to ensure that the question is meaningful for health care decision-making. Assembling a multi-disciplinary guideline development group in which relevant stakeholders are represented comprises the first step. The panel is responsible for deciding on the goal(s) and scope of the guideline, by formulating priority questions. The panel then translates priorities into an answerable PICO research question format. Respondents described how preselecting topics that fit within the PICO format frames questions in light of available evidence.

[‡]These items assess whether guidelines provide suggestions on how to organize and integrate care services efficiently in response to changes in demand for health care.

[§]Collaboration agreements may be addressed elsewhere in separate documents on regional care coordination.



Stakeholder involvement

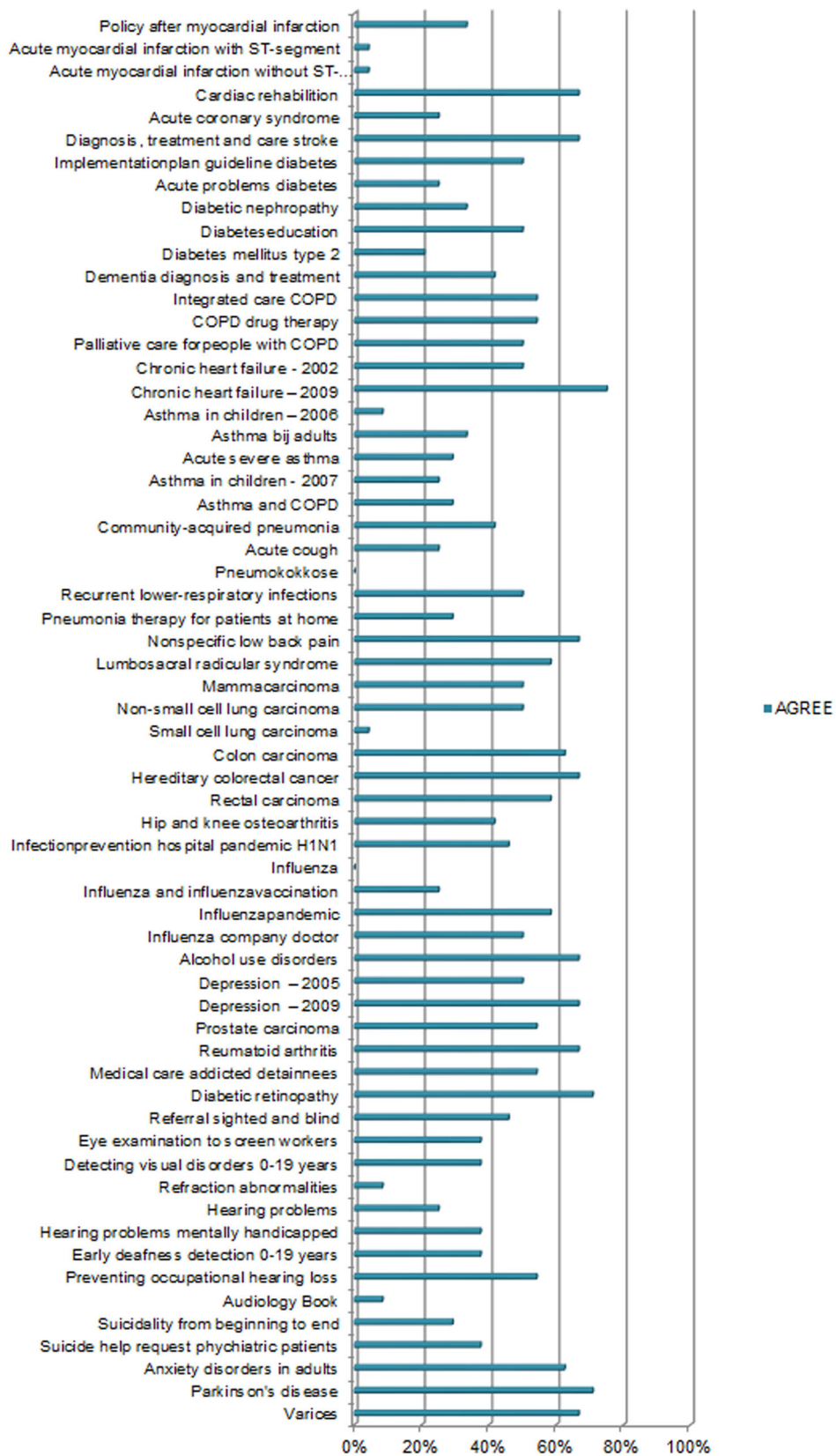


FIGURE 6 AGREE scores on stakeholder involvement



Applicability

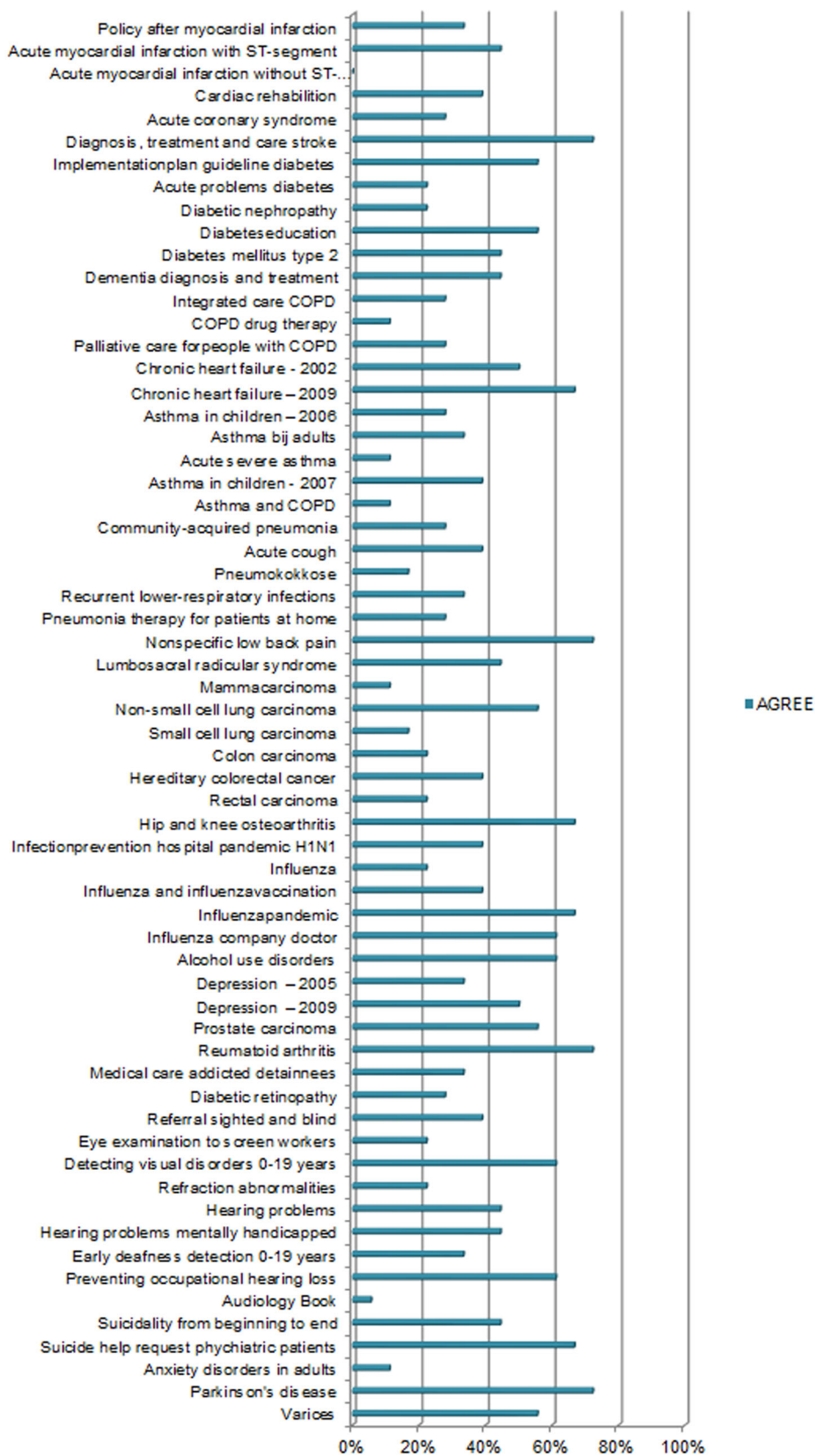


FIGURE 7 AGREE scores on applicability



Patient involvement

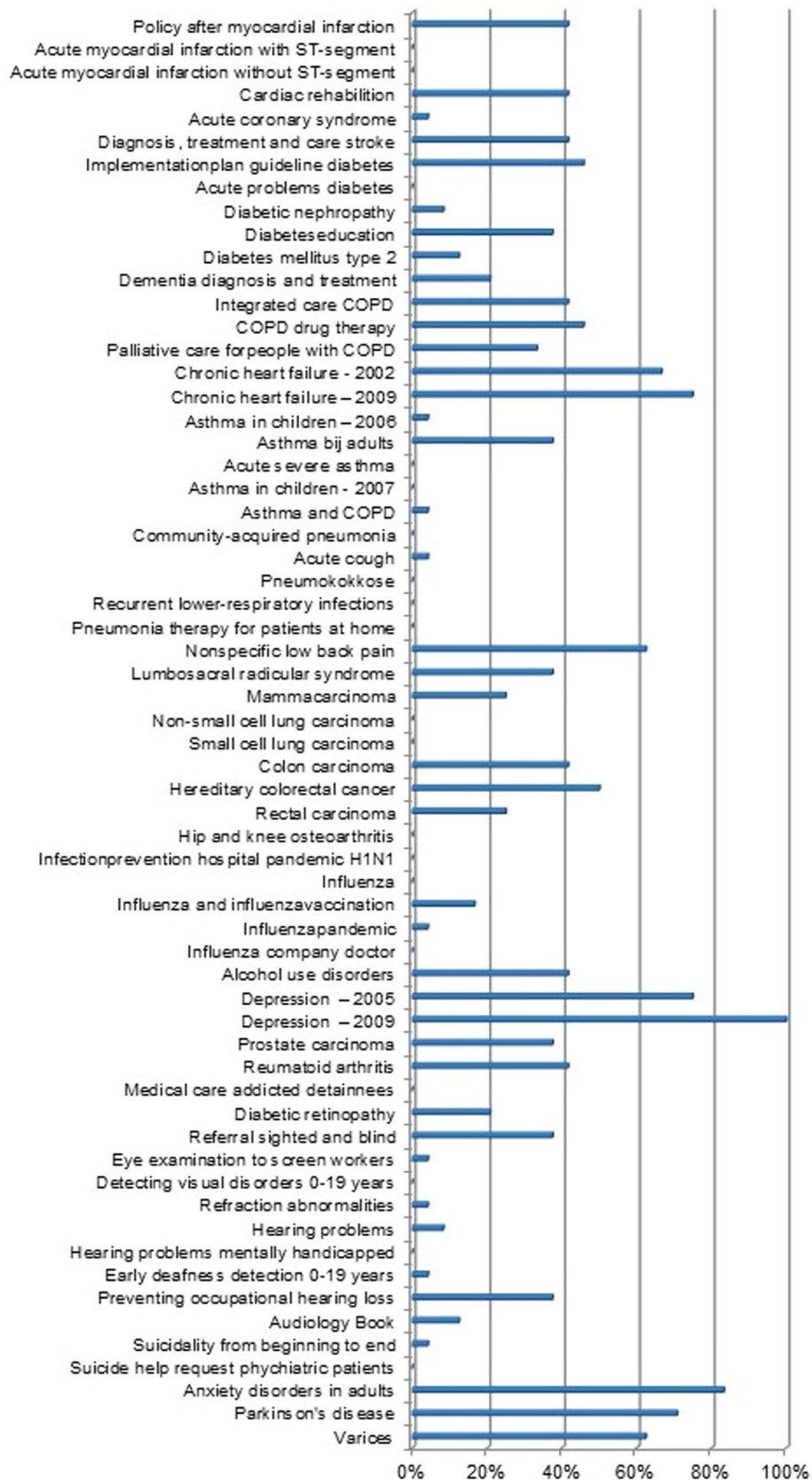


FIGURE 8 Additional items on patient involvement



User involvement

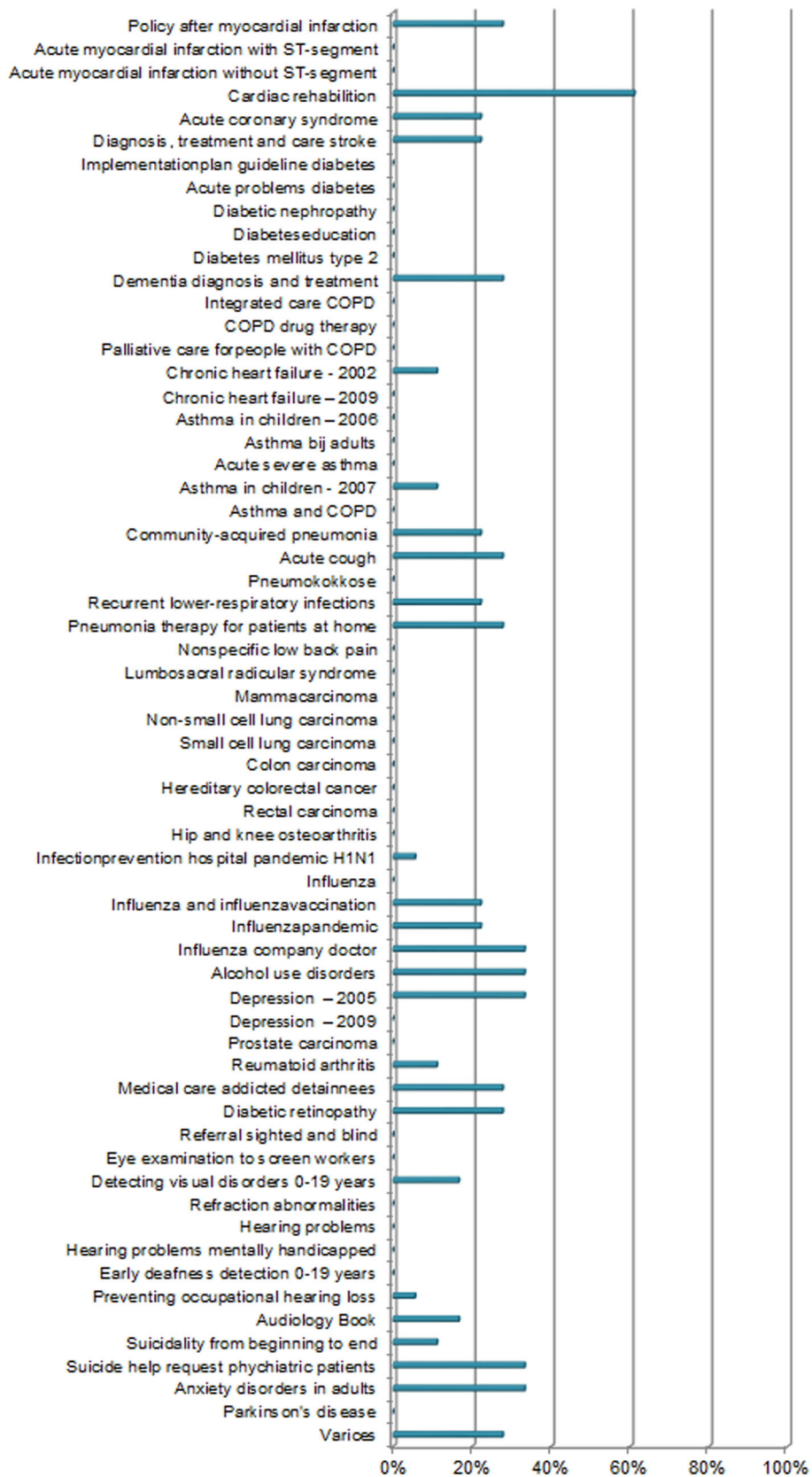


FIGURE 9 Additional items on user involvement

Healthcare organisation

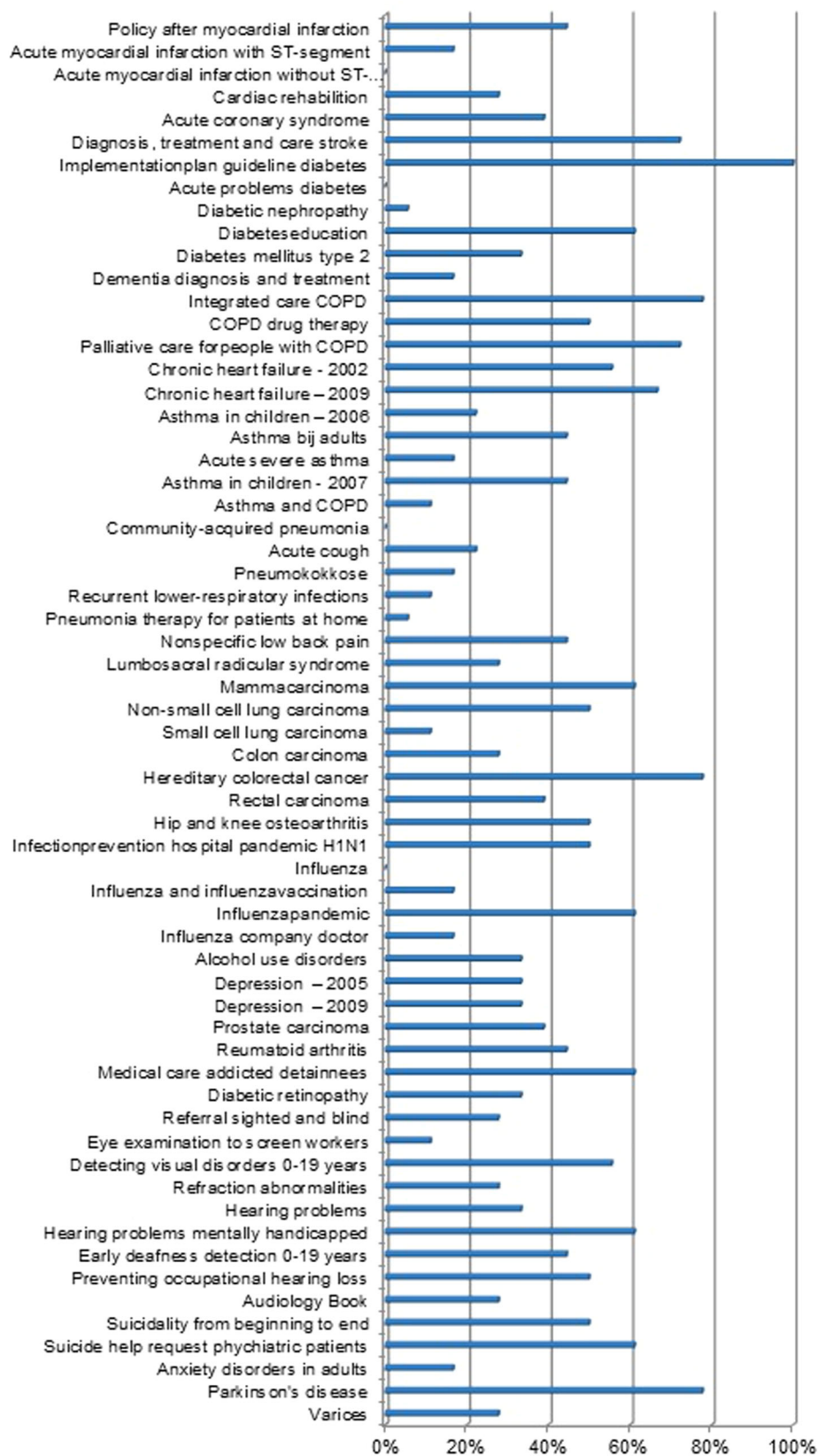


FIGURE 10 Additional items on the practice of health care organization



Cost implications

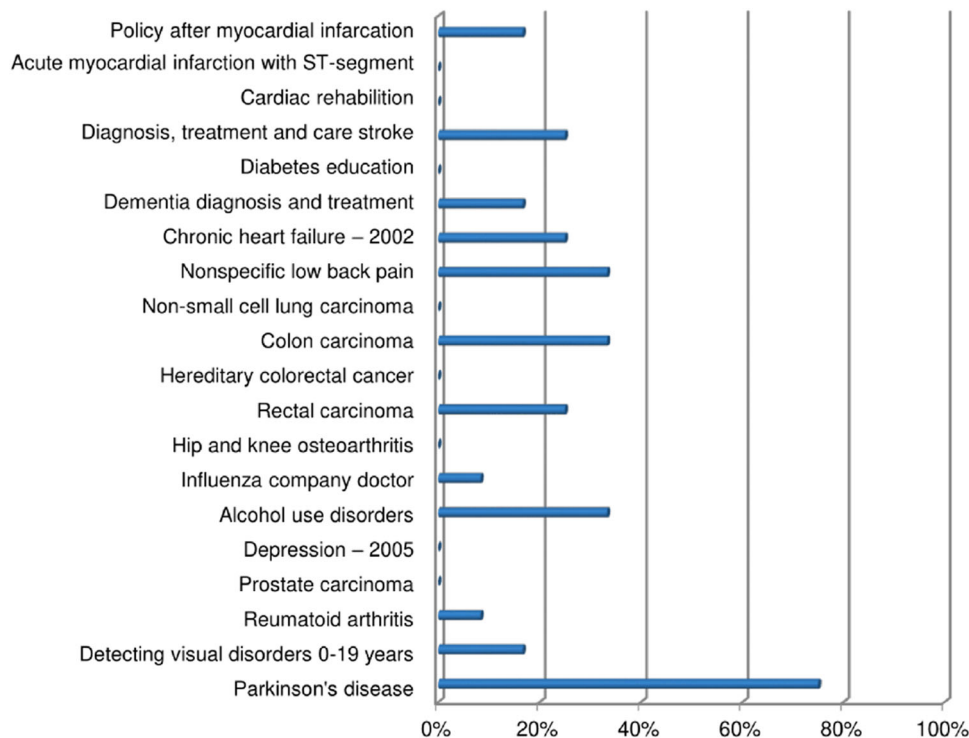


FIGURE 11 Additional items on cost implications

Everybody was so focused on the literature; it just went beyond my personal expertise. It was too big a difference [for me to meaningfully participate in those discussions] (GP)

Such procedures tend to overlook the reasons behind gaps in scientific literature. Several respondents described how aspects critical to providing good quality care remain unresearched because there is no financial interest or there are ethical concerns for carrying out such study designs.

Much of the evidence is pharmaceutical-biased, a selection of topics that serves the interest of pharmaceutical companies (.) [and] there are simply some topics that will never be studied scientifically because you cannot ethically justify a control group (physician, chairperson guideline).

3.3 | Collecting evidence

After the questions have been defined, evidence is systematically collected and reviewed. Respondents described that this part of the development process is mostly the domain of independent methodologists, shifting evidence synthesis from the guideline panel to an independent expert or institution. In the quote below, a respondent

clarifies that health professionals are deliberately excluded to avoid conflict of interest.

It is funny how far methodological institutes are removed from the realities of clinical practice. Honestly, it is quite cumbersome to let them do the research, but there is just no other way. If [clinicians] would do the research themselves, the evidence would automatically be considered biased (physician, chairperson guideline).

Although many respondents acknowledge that specific methodological expertise is needed to conduct such evidence reviews, clinical expertise is considered equally important to avoid a 'naive understanding of science'. One respondent clarifies:

Methodologists only look at the study design and the study findings (.) that is why [our standard for guideline development] states that evidence summaries must be written by methodologists AND clinicians. More practitioners need to be trained in methodology (physician, department chairperson Cochrane Collaboration).

In spite of such recommendations in standards for guideline development, the involvement of clinicians in this part of the guideline development process is only structurally embedded in some exceptional



Safety

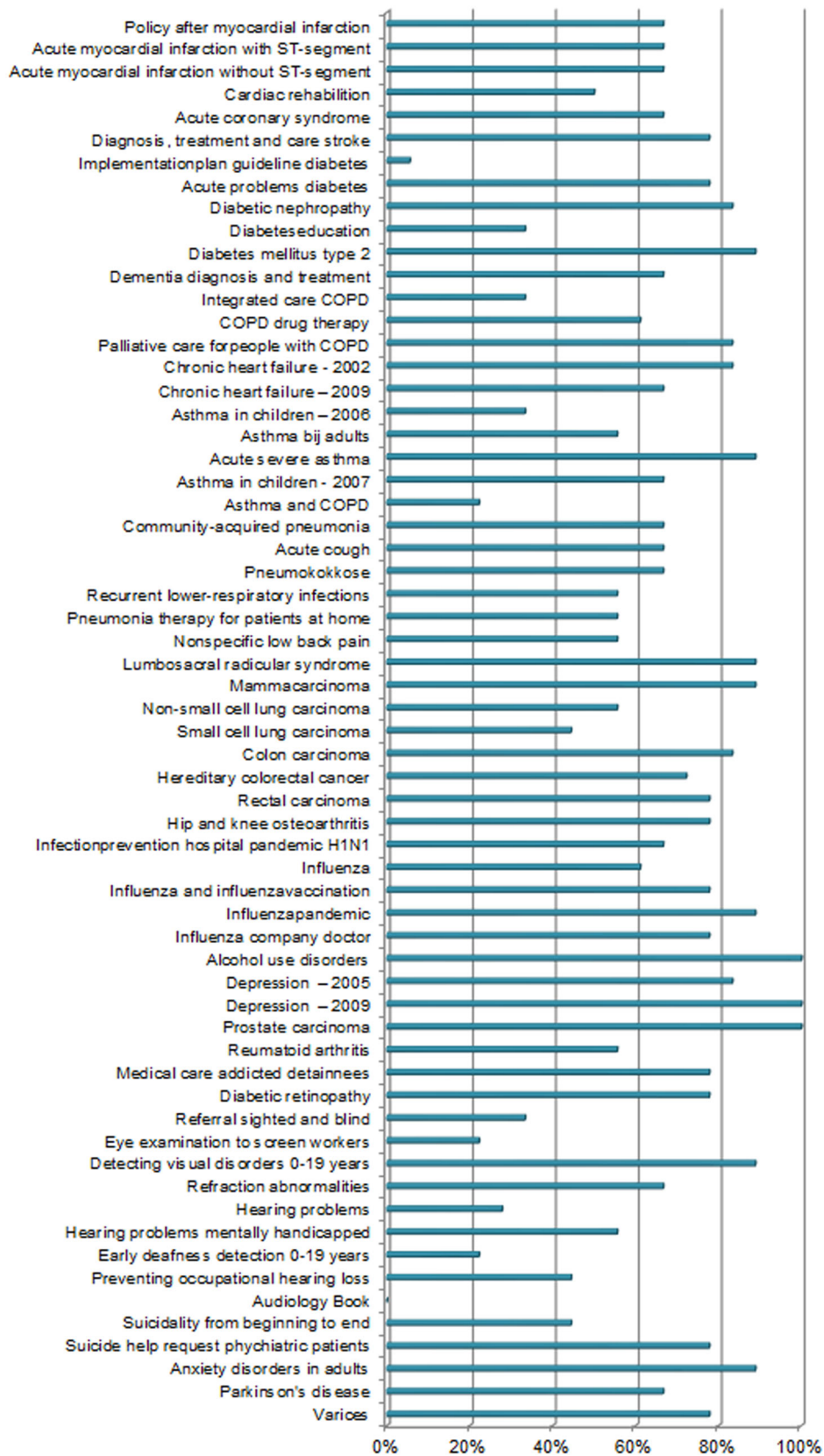


FIGURE 12 Additional items on safety



cases such as the Dutch Association of Primary Care Physicians (NHG) that requires their developers and methodologists to practice general medicine part-time. Many respondents in other domains voiced their concerns about the *external* validity of the evidence.

3.3.1 | Integrating evidence

After the evidence is collected, findings are integrated. Figure 1 suggests that evidence profiles are informed by the *outcomes across studies*. Some respondents mentioned, however, that some topics simply lack the supporting scientific evidence base, thus requiring consensus.

Especially for health care organisations, there is absolutely no scientific evidence because you cannot standardize the complexity of clinical practice in a trial. [That is why this topic is addressed based on] the opinion of the workgroup panels complemented by expert opinion papers (secretary guideline).

Nonetheless, respondents overall agreed that the balance has shifted from consensus-based to predominantly science-based. A respondent who has been participating in guideline development from the beginning notes:

The first guidance was largely based on consensus, experienced-based knowledge complemented with scientific literature. Over time, however, it has gone in quite a different direction, the weight has shifted largely in favour of scientific literature (GP).

This by no means implies that respondents are naïve about the risks of consensus. A physician and department chairperson of the Cochrane Collaboration notes that power dynamics may skew the outcomes: *'I do not think that consensus is necessarily a good replacement because the outcome of consensus is strongly affected by the panel composition, the facilitator and how the consensus meeting was set up'*. What is considered admissible evidence thus seems strongly related to the reliability that can be ensured through procedures followed, which seems to have made consensus an unsought basis for guidance. Even the RAND Delphi method, the only procedure that GRADE acknowledges as a robust consensus method, has not been used in any of the 62 guidelines that we evaluated.

3.3.2 | Evidence conclusion

To draw up conclusions based on evidence, the quality of all outcomes is appraised. Tools are used to structure transparent and explicit reporting of steps to appraise the quality of the evidence. Reflecting upon the widespread adoption of GRADE within the guideline community internationally, respondents indicate that this adoption seemed based on the methodological improvements through the use of GRADE, to

improve the transparent and systematic use of criteria to assess the quality of the evidence, compared to previous methods. One respondent described this shift as indicative of the 'scientization' of guideline development, by overemphasizing best external evidence, which he remarked is but one of the three facets of the philosophy of EBM. Others mentioned that developers and panel members increasingly approach guidelines the same way they would scientific articles: focusing only on the conclusion.

Sackett and colleagues proposed this trinity of the best available evidence, clinical expertise, and patient preferences, but to a lot of guideline users, developers and others involved it is about the evidence which means that a lot of people stop there. (...) Guideline panels feel that guideline's conclusions that are supported by strong evidence do not require [scrutiny of the expert panel], but can directly be translated into recommendations (project leader guideline).

The assumption that the gold standard of 'strong evidence', that is systematic reviews of RCTs, generates certainty that is relevant for quality improvement, was considered problematic by many respondents. As a respondent concluded, the scientization of CPGs and the inherent drawbacks from focusing exclusively on scientific evidence may delegate critical dilemmas to the consulting room:

The risk is that clinicians think decision-making should be limited to evidence-based medicine. [but] there are simply some topics that will never be studied scientifically (physician, chairperson guideline).

This concern was related to the education of medical doctors, as several respondents noted that the current focus on EBM in such education makes young clinicians wonder whether they should refrain from acting when there is no strong evidence at hand to guide decision-making.

3.3.3 | Formulating recommendations

In the previous sections, we described that CPGs that are interpreted as solely evidence-based can only provide guidance for clinical situations supported by strong scientific evidence. One respondent describes that questions that cannot be researched in an RCT design or measured quantitatively, such as patient preferences and applicability, are more difficult to appraise using GRADE. As depicted in Figure 1, such considerations may be included once evidence summaries have been completed. Despite attempts to formalize practical considerations in guideline development, the quantitative analysis shows significantly lower AGREE scores on domains, including applicability and stakeholder involvement, that lack this type of supporting evidence. This might be explained by the difficulty that many

respondents described in aligning the different types of knowledge derived from science and from experience.** Moreover, two respondents remarked that there is no official structure to guide the process of connecting research evidence and other considerations, nor should these be connected.

Collecting the evidence and translating it to the national context using the other considerations: 'what do patients value?' or 'what do our professionals think is important' (...) that is not science. That is about opinions and experience (...) [You connect the evidence to those considerations] in your final recommendations. You start from the scientific literature, you summarise findings in the conclusion and together with the other considerations you formulate recommendations (...) Linking those two, I would not do that, because you would get an undesirable mixture of science and experience (guideline development expert).

Like the quote suggests, separating evidence and experience is the dominant mode developers use for bypassing this issue, by clearly separating the process of evidence summary and formulating recommendations. This is considered one of the main advantages of the Evidence-to-decisions framework outlined by GRADE. This separation also allows for distinguishing between recommendations based on expert opinion and recommendations based on science. However, as the quantitative results of our study show, this in fact often leads to an underrepresentation of recommendations that are important for quality improvement but for which the research evidence is weak. Respondents described how such topics often end up in separate documents and appendices. They described this phenomenon as strategies that are used to bridge the distance between 'what we expect a guideline to be' (evidence-based), and 'what end-users actually need'.

Basically, [part 1] of the guideline is how you would expect a guideline to be. In addition, we have a section on health care organisation [part 3] which is actually the part that end-users are interested in (secretary guideline).

Alternatively, methods for collecting user feedback could be used to bring CPGs closer to the needs of end-users. Such methods are likely to help to test and contextualize guideline recommendations in practice. However, the quantitative analysis shows scarce reporting of such methods (<10% of guidelines included a pilot test). The few respondents that reported collecting user feedback critically noted that the feedback almost exclusively remarks on the literature that supports the recommendations and the phrasing of the

recommendations, rather than the applicability and clinical relevance of the guideline's recommendations.

We do a pilot to examine whether the guideline is implementable because we receive very few responses to our requests for feedback. The ones that do respond just say: 'this is fine, I can work with this' or only focus on the phrasing and the literature that supports the recommendations. [But this is not why we collect feedback from professionals] we want to examine whether the guideline is feasible! (guideline development expert)

The fact that guideline panels include health care providers neither fills this need for feasibility knowledge. Practical considerations of feasibility and testing whether recommendations sufficiently contribute to the quality issues professionals experience are thus underrepresented in guideline development.

3.4 | What is a CPG (for)?

Findings show that the system of guideline development puts emphasis on the rigour of methodologies used for grading evidence, whereas rigorous methods for formulating recommendations that are unable to draw on such an evidence base, such as consensus methods have dissipated. Efforts to bypass tensions between scientific evidence and 'other' evidence inadvertently lower the visibility of quality aspects that lack a strong scientific evidence base.

Respondents acknowledge the value of the dominant system of guideline development for questions with a clear-cut and strong evidence base. When knowledge is more controversial and ambiguous, however, they feel ill-equipped to formulate a recommendation. There are other approaches that can be used to formulate recommendations, without following the currently dominant evidence-based method to the letter, but these have lost legitimacy and methodological infrastructure. The respondent quoted below calls for a more flexible approach to guideline development and encourages a critical reflection on the usefulness of the currently dominant evidence-based method, especially for topics that are not, nor could be supported by this one type of strong evidence.

We have to make clearer distinctions between the topics for which evidence-based guideline development works and the topics for which it does not (.) there is too little scientific evidence about health care organisation, so (.) we organised a very good focus group with patients and professionals (...) We need much more flexibility to allow [such topics to be properly addressed] (.) This more creative way of thinking [about guidelines] is, however, rarely used. It is either about quality of care or it is a guideline. Few people manage to combine those two (physician, department chairperson Cochrane Collaboration).

**Although this refers to the experiences of all of the guidelines' target users, our quantitative assessment of the level of *patient involvement* in evaluated CPGs indicates that the relative weight of patients' perspectives remains an important point for discussion.



This notion of considering what type of knowledge fits best with addressing the quality concern at hand, rather than what type of knowledge fits best with the dominant guideline development system, resonates with responses from other respondents. Instead of guideline panels following the steps laid out by the system, the system would need to adapt to the questions that are laid out by the guideline panel.

4 | DISCUSSION

This study examined guideline quality in relation to the availability of certain types of evidence and reflects on the implications of CPGs' promise to improve the quality of care practices. The quantitative analysis suggests that rigorous methods for evidence appraisal take precedence over other considerations, although it has been acknowledged since the early days of EBM that such other considerations may be crucial to avoid that CPGs becoming 'evidence-biased'.³⁶ Respondents marked this emphasis on methodological rigour as the 'scientization' of guideline development, which lowers the inclusion of quality aspects that lack a strong scientific evidence base.

4.1 | Limitations

This article reports on an evaluation of Dutch CPGs that was conducted in 2010. We believe our findings continue to provide a relevant contribution to ongoing discussions and concerns regarding the variable quality of CPGs that may limit the uptake, utilization, and overall efficacy of CPGs to improve care practices.⁶ Studies indicate that guideline development methodologies like GRADE have generated improvements in the rigour with which CPGs follow standardized development methods and the reporting of this process.^{6,37,38} Quoting Armstrong and colleagues⁶ 'improvement is evident, but still necessary in CPG quality', particularly with regard to *applicability*. A review of 421 CPGs for the management of noncommunicable diseases in primary care found that even high-quality guidelines, that is, guidelines that scored high on 'rigour of development', performed poorly on *applicability* to practice,³⁸ which suggests that standardized frameworks to promote the methodological quality of searching, selecting and presenting evidence in guideline development have an indirect relationship to the quality of the guidance itself. Norris and Bero³⁹ note that GRADE is currently not applicable to many quality questions that guideline developers face, especially quality questions that require developers to draw on different types of knowledge. Guideline developers particularly struggled with this problem during the Covid-19 pandemic, where guidance had to be developed under extreme time pressure and knowledge uncertainty. A mixed-methods study⁴⁰ has highlighted the pertinence of developing methods to produce credible guidance in the absence of 'high-quality' evidence. It is expected that the Covid-19 pandemic may act as a catalyst for addressing a problem that has existed long before the pandemic: the scientific uncertainties that are always part of guideline development and require developers to balance scientific robustness with feasibility, acceptability,

adequacy, and contingency.^{40–42} In spite of this study's limitations, we feel that our study's contribution to the current state of guideline development remains relevant. In the subsequent paragraphs, we outline how this study contributes to the question of how the efficacy of CPGs as tools for quality improvement can be strengthened.

4.1.1 | Appraising and including different types of knowledge

Initiatives like the GRADE 'evidence to decision framework' show laudable attempts to formalize practical considerations that affect decision-making processes in guideline development.²⁷ But additional steps seem required for CPGs to maintain a focus on quality improvement. Building on this study's findings, we formulate recommendations to support this move towards integrating evidence and quality considerations, thus bringing CPGs closer to their promise of serving as quality instruments. Respondents suggested using practice questions as a guide for identifying the most fitting evidence for quality concerns. Although such questions may cover a range of topics, manuals on guideline development continue to frame practice questions into a PICO question format,³ which only accommodates therapy questions.⁴³ Given that questions about the effectiveness of therapy are indeed best served by (meta-reviews of) scientific evidence from RCTs,⁴⁴ this is an important part of funneling what gets to count as relevant knowledge in guideline development. It follows that for CPGs to serve as quality instruments, a variety of question formats are needed to accommodate different types of quality questions. To address questions that lack a strong scientific evidence base, such as those concerning health care organization and cost-effectiveness, developers may appeal to other sources of knowledge such as quality improvement, expert opinion, input from patients and professionals, and best practice information. This has always been part of the discourse of CPGs, but as our quantitative analysis shows, not of its practices. We recommend expanding the second development phase, traditionally coined 'evidence synthesis' that gives primacy to knowledge derived from research to 'knowledge synthesis' that resonates with the notion of appraising and including 'most fitting evidence' to answer quality questions from practice stretching well beyond the PICO format. To support the reliable inclusion of diverse sources of knowledge, further research is needed to develop methods for the robust inclusion of other types of knowledge that need to be evaluated drawing on a broader range of modes of scientific reasoning.³¹

4.1.2 | Explicating the goal of clinical practice guidelines

We recommend developers to be more explicit about how end-users should apply CPGs. Notions of 'bottom-up EBM', put forward by the founders of EBM to avoid allegations of 'cookbook medicine',⁴⁴ still leave the role of CPGs sufficiently vague.⁴⁵ Weaver remarks: 'EBM's incompleteness lies in having inadequately articulated where the

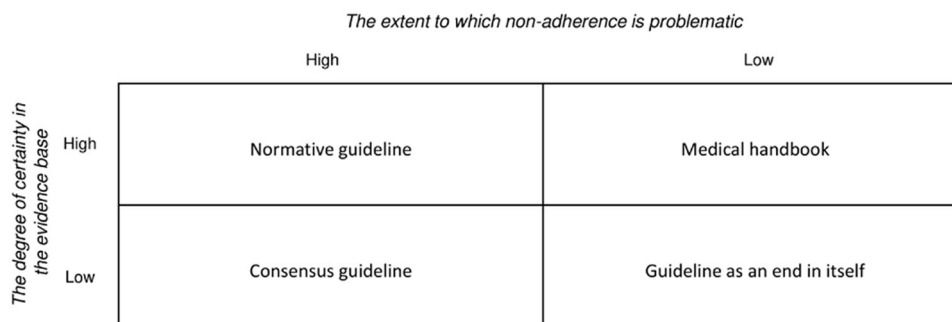


FIGURE 13 Four types of CPGs

evidence of EBM belongs in the clinical care process'.⁴⁵ Besides diversity in patient preferences, practitioners choose specific treatments that, based on clinical expertise, are considered crucial to provide high-quality care. However, other rationales including habit and partial interests may also underlie such decisions.⁴⁶ Although epidemiological studies highlight safety risks that accompany practice variation and low guideline adherence,^{47–49} other studies highlight important reasons for guideline divergence to ensure safe care.^{50–56} It follows that to clarify how CPGs inform clinical work, is to explicate whether nonadherence is problematic or not. To facilitate such reflection, Figure 13 proposes a framework that distinguishes four types of CPGs.

1. Normative guideline: guideline adherence is considered the norm to deliver safe, high-quality care. Cases of guideline divergence need to be reported and are only allowed, provided that there are legitimate reasons to do so.
2. Consensus guideline: for topics for which guidance is needed (e.g., because there is little agreement on good practice), yet lack a solid scientific evidence base, developers may appeal to other sources of knowledge to provide answers to pressing quality questions. Monitoring adherence is recommended. Cases of guideline divergence need to be reported and are only allowed, provided that there are legitimate reasons to do so.
3. Medical handbook: when there is agreement on relevant knowledge among practitioners, a guideline would serve only as a handbook. Monitoring adherence would be futile, as there are plenty of reasons why divergence may be needed to deliver safe, high-quality care. Required reports of divergence would only serve to bureaucratize health care rather than safeguard its quality.
4. Guideline is an end in itself: when practice variation does not affect safety or quality of care, guidelines would become an end in itself. Monitoring adherence would be squandering resources.

We propose this framework be used as a reflective tool to ascertain whether guidelines actually serve as instruments for quality (Types 1 and 2): the other types (3 and 4) serve as heuristic guidance for guideline forms that should be avoided. Focusing on the most pertinent quality questions for patient care resonates with the reported need to move away from classic textbook-like guidelines.⁸ A

focus on the goal rather than method treats the evidence as in service of quality, instead of the other way around.

5 | CONCLUSION

This study is the first to empirically show the consequences of dominant systems of guideline development and reflects on what is needed for CPGs to serve as instruments for quality. Although the dominant system works well for questions with a clear-cut and strong scientific evidence base, it poorly accommodates topics that urgently require guidance, but that lack such a base. Findings highlight the need to scope guidelines with a strong focus on the most pertinent quality issues that point developers towards the most fitting evidence for the question at hand, thus treating evidence as being in service of quality.

ACKNOWLEDGEMENTS

The authors are grateful for contributions in study design and data collection by Siok Swan Tan, Leona Hakkaart-van Roijen, and Werner Brouwer. We also greatly appreciate the time spent by guidance developers and guideline development experts to participate in this study. Finally, we would like to thank the Dutch Council for Quality of Health care, especially Dunja Dreesens en Jannes van Everdingen, for support in the design of this study. This study was funded by the Dutch Council for Quality of Health care, The Hague, The Netherlands.

CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interest.

ETHICS STATEMENT

Following Dutch ethics regulations, no ethics approval was needed to conduct interviews with professionals. The Ethics Review Committee of the Erasmus School of Health Policy and Management of the Erasmus University Rotterdam marked this study as not requiring further evaluation by the Research Ethics Review Committee.

AUTHOR CONTRIBUTIONS

Teun Zuiderent-Jerak, Sonja Jerak-Zuiderent, Hester van de Bovenkamp and Roland Bal designed the study methodology. Data was collected by Teun Zuiderent-Jerak, Sonja Jerak-Zuiderent and Hester van de



Bovenkamp. All authors reviewed emerging themes and categories during data analysis and contributed comments to aid reflexivity. Marjolein Moleman and Teun Zuiderent-Jerak drafted the initial manuscript with all authors contributing significantly to revising this for submission. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The quantitative data are available from the corresponding author on reasonable request. Due to the nature of this study, interview participants of this study did not agree for their data to be shared publicly, so supporting qualitative data are not available.

ORCID

Marjolein Moleman  <http://orcid.org/0000-0002-5289-2341>

Sonja Jerak-Zuiderent  <https://orcid.org/0000-0003-2804-1763>

Hester van de Bovenkamp  <https://orcid.org/0000-0002-6412-2707>

Teun Zuiderent-Jerak  <https://orcid.org/0000-0003-3008-1297>

REFERENCES

- Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet*. 2000;355(9198):103-106. doi:10.1016/S0140-6736(99)02171-6
- Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E. *Clinical Practice Guidelines We Can Trust*. National Academy of Sciences; 2011. http://www.nap.edu/catalog.php?record_id=13058%0Ahttps://www.awmf.org/fileadmin/user_upload/Leitlinien/International/IOM_CPG_lang_2011.pdf
- Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. GRADE Working Group. Published October 2013. Accessed May 3, 2021. <https://gdt.gradepro.org/app/handbook/handbook.html>
- Kredo T, Bernhardsson S, Machingaidze S, et al. Guide to clinical practice guidelines: the current state of play. *Int J Qual Heal Care*. 2016;28(1):122-128. doi:10.1093/intqhc/mzv115
- Burda BU, Norris SL, Holmer HK, Ogden LA, Smith MEB. Quality varies across clinical practice guidelines for mammography screening in women aged 40-49 years as assessed by AGREE and AMSTAR instruments. *J Clin Epidemiol*. 2011;64(9):968-976. doi:10.1016/j.jclinepi.2010.12.005
- Armstrong JJ, Goldfarb AM, Instrum RS, MacDermid JC. Improvement evident but still necessary in clinical practice guideline quality: a systematic review. *J Clin Epidemiol*. 2017;81:13-21. doi:10.1016/j.jclinepi.2016.08.005
- Burgers JS. Guideline quality and guideline content: are they related? *Clin Chem*. 2006;52(1):3-4. doi:10.1373/clinchem.2005.059345
- Alonso-Coello P, Irfan A, Solà I, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Qual Saf Heal Care*. 2010;19(6):1-7. doi:10.1136/qshc.2010.042077
- Ciquier G, Azzi M, Hébert C, Watkins-Martin K, Drapeau M. Assessing the quality of seven clinical practice guidelines from four professional regulatory bodies in Quebec: what's the verdict. *J Eval Clin Pract*. 2020;27:25-33. doi:10.1111/jep.13374
- Knaapen L. Being 'evidence-based' in the absence of evidence: the management of non-evidence in guideline development. *Social Stud Sci*. 2013;43(5):681-706. doi:10.1177/0306312713483679
- Moreira T. Diversity in clinical guidelines: the role of repertoires of evaluation. *Soc Sci Med*. 2005;60(9):1975-1985. doi:10.1016/j.socscimed.2004.08.062
- Knaapen L, Cazeneuve H, Cambrosio A, Castel P, Fervers B. Pragmatic evidence and textual arrangements: a case study of French clinical cancer guidelines. *Soc Sci Med*. 2010;71(4):685-692. doi:10.1016/J.SOCSCIMED.2010.05.019
- Busse R, Klazinga N, Panteli D, Quentin W. *Improving Healthcare Quality in Europe*. European Observatory on Health Systems and Policies; 2019.
- Brouwers MC, Spithoff K, Kerkvliet K, et al. Development and validation of a tool to assess the quality of clinical practice guideline recommendations. *JAMA Netw Open*. 2020;3(5):e205535. doi:10.1001/jamanetworkopen.2020.5535
- Rosenfeld RM, Wyer PC. Stakeholder-driven quality improvement: a compelling force for clinical practice guidelines. *Otolaryngol Neck Surg*. 2018;158(1):16-20. doi:10.1177/0194599817735500
- Kowalski CJ, Mrdjenovich AJ, Redman RW. Scientism recognizes evidence only of the quantitative/general variety. *J Eval Clin Pract*. 2020;26(2):452-457. doi:10.1111/jep.13330
- Tonelli MR, Upshur REG. A philosophical approach to addressing uncertainty in medical education. *Acad Med*. 2019;94(4):507-511. doi:10.1097/ACM.0000000000002512
- Falzer PR. Evidence-based medicine's curious path: from clinical epidemiology to patient-centered care through decision analysis. *J Eval Clin Pract*. 2020;July:1-7. doi:10.1111/jep.13466
- Blunt C. Hierarchies of evidence in evidence-based medicine. Hierarchies Evid Evidence-based Med. 2015. <http://theses.lse.ac.uk/id/eprint/3284>
- Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):A12. doi:10.7326/ACPJC-1995-123-3-A12
- Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:4008. doi:10.1136/bmj.j4008
- Whiting P, Savovic J, Higgins JPT, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-234. doi:10.1016/J.JCLINEPI.2015.06.005
- Schünemann H, Brozek J, Guyatt G, Oxman A. *GRADE Handbook*. Published 2013. Accessed January 10, 2021. <https://gdt.gradepro.org/app/handbook/handbook.html#h.9rdbelsnu4iy>
- Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is "quality of evidence" and why is it important to clinicians. *BMJ*. 2008;336(may):995-998. doi:10.1136/bmj.39490.551019.be
- Siemieniuk R, Gordon G. *What is GRADE? | BMJ Best Practice*. BMJ Best Practice. Published 2018. Accessed January 10, 2021. <https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/>
- Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. *Heal Res Policy Syst*. 2006;4:1-6. doi:10.1186/1478-4505-4-13
- Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 2016;353:353. doi:10.1136/bmj.i2016
- Moberg J, Oxman AD, Rosenbaum S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Heal Res Policy Syst*. 2018;16(1):45. doi:10.1186/s12961-018-0320-2
- Mercuri M, Gafni A. The evolution of GRADE (part 3): a framework built on science or faith? *J Eval Clin Pract*. 2018;24(5):1223-1231. doi:10.1111/jep.13016
- Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: clinical practice guidelines. *BMJ*. 2016;353:353. doi:10.1136/BMJ.i2089

31. Wieringa S, Dreesens D, Forland F, et al. Different knowledge, different styles of reasoning: a challenge for guideline development. *BMJ evidence-based Med.* 2018;23(3):87-91. doi:10.1136/bmjebm-2017-110844
32. Agree C. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Heal Care.* 2003;12(1):18-23. doi:10.1136/qhc.12.1.18
33. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting, and evaluation in health care. *Prev Med.* 2010;51(5):421-424. doi:10.1016/j.ypmed.2010.08.005
34. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res.* 2005;15(9):1277-1288. doi:10.1177/1049732305276687
35. Nuckols TK, Shetty K, Raaen L, Khodyakov D. Technical quality and clinical acceptability of a utilization review guideline for occupational conditions OGD[®] Treatment Guidelines by the Work Loss Data Institute Santa Monica; 2017. https://www.rand.org/pubs/research_reports/RR1819.html
36. Evans JG. Evidence-based and evidence-biased medicine. *Age Ageing.* 1995;24(6):461-463. doi:10.1093/ageing/24.6.461
37. Lunny C, Ramasubbu C, Puil L, et al. Over half of clinical practice guidelines use non-systematic methods to inform recommendations: a methods study. *PLoS One.* 2021;16:0250356. doi:10.1371/journal.pone.0250356
38. Molino CGRC, Leite-Santos NC, Gabriel FC, et al. Factors associated with high-quality guidelines for the pharmacologic management of chronic diseases in primary care: a systematic review. *JAMA Intern Med.* 2019;179(4):553-560. doi:10.1001/jamainternmed.2018.7529
39. Norris SL, Bero L. GRADE methods for guideline development: time to evolve? *Ann Intern Med.* 2016;165(11):810-811. doi:10.7326/M16-1254
40. Moleman M, Macbeth F, Wieringa S, et al. From "getting things right" to "getting things right now": developing COVID-19 guidance under time pressure and knowledge uncertainty. *J Eval Clin Pract.* 2022;28(1):49-56. doi:10.1111/jep.13625
41. Shaw RL, Larkin M, Flowers P. Expanding the evidence within evidence-based healthcare: thinking about the context, acceptability and feasibility of interventions. *BMJ evidence-based Med.* 2014;19(6):201-203. doi:10.1136/eb-2014-101791
42. Browman GP, Somerfield MR, Lyman GH, Brouwers MC. When is good, good enough? Methodological pragmatism for sustainable guideline development. *Implement Sci.* 2015;10:28. doi:10.1186/s13012-015-0222-4
43. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Serv Res.* 2014;14(1):529. doi:10.1186/s12913-014-0579-0
44. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *Br Med J.* 1996;312(7023):71-72. Accessed March 7, 2020. <https://www.jstor.org/stable/29730277>
45. Weaver RR. Reconciling evidence-based medicine and patient-centred care: defining evidence-based inputs to patient-centred decisions. *J Eval Clin Pract.* 2015;21(6):1076-1080. doi:10.1111/jep.12465
46. Light DW. Health-care professions markets and countervailing powers. *Book of Medical Sociology.* Vanderbilt University Press; 2000. https://www.researchgate.net/publication/280798828_Health-care_professions_markets_and_countervailing_powers
47. Grilli R, Lomas J. Evaluating the message: the relationship between compliance rate and the subject of a practice guideline. *Med Care.* 1994;32(3):202-213.
48. Burstin HR, Conn A, Setnik G, et al. Benchmarking and quality improvement: the Harvard Emergency Department Quality Study. *Am J Med.* 1999;107(5):437-449. doi:10.1016/S0002-9343(99)00269-7
49. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines?: a framework for improvement. *J Am Med Assoc.* 1999;282(15):1458-1465. doi:10.1001/jama.282.15.1458
50. Boyd CM, Darer J, Boulton C, Fried LP, Boulton L, Wu AW. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. *J Am Med Assoc.* 2005;294(6):716-724. doi:10.1001/jama.294.6.716
51. Kumar A, Smith B, Pisanelli DM, Gangemi A, Stefanelli M. Clinical guidelines as plans: an ontological theory. *Methods Inf Med.* 2006;77:204-210. doi:10.1055/s-0038-1634052
52. Greer AL, Goodwin JS, Freeman JL, Wu ZH. Bringing the patient back in: guidelines, practice variations, and the social context of medical practice. *Int J Technol Assess Health Care.* 2002;18(4):747-761. doi:10.1017/S0266462302000569
53. Oertle M, Bal R. Understanding non-adherence in chronic heart failure: a mixed-method case study. *BMJ Qual Saf.* 2010;19(6):e37. doi:10.1136/qshc.2009.033563
54. Walter LC, Davidowitz NP, Heineken PA, Covinsky KE. Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA performance measure. *J Am Med Assoc.* 2004;291(20):2466-2470. doi:10.1001/jama.291.20.2466
55. Van Weel C, Schellevis FG. Comorbidity and guidelines: conflicting interests. *Lancet.* 2006;367(9510):550-551. doi:10.1016/S0140-6736(06)68198-1
56. Weingarten S. Translating practice guidelines into patient care: guidelines at the bedside. *Chest.* 2000;118(2 SUPPL.):4S-7S. doi:10.1378/chest.118.2_suppl.4S

How to cite this article: Moleman M, Jerak-Zuiderent S, van de Bovenkamp H, Bal R, Zuiderent-Jerak T. Evidence-basing for quality improvement; bringing clinical practice guidelines closer to their promise of improving care practices. *J Eval Clin Pract.* 2022;28:1003-1026. doi:10.1111/jep.13659