# Articles

## Systematic quality review of clinical guidelines – feasible and useful?

**Anne Hilde Røsvik**
*Project Manager, Norwegian Electronic Health Library, Norwegian Knowledge Centre for the Health Services (NOKC), PO Box 7004 St Olavs plass, N-0130 Oslo, Norway; aro@helsebiblioteket.no*
**Espen Movik**
*Researcher in Health Economics, Norwegian Knowledge Centre for the Health Services (NOKC), Oslo, Norway*
**Magne Nylenna**
*Editor, Norwegian Electronic Health Library, Norwegian Knowledge Centre for the Health Services; Professor of Community Medicine, University of Oslo and Norwegian University of Science and Technology, Trondheim, Norway*

**Abstract**
**Background**: Many clinical guidelines are produced and used, but there is no consensus on how to evaluate their quality. The purpose of this study was to test the feasibility and the benefit of systematic quality review of clinical guidelines.
**Methods**: 127 Norwegian guidelines were evaluated with the AGREE (Appraisal of Guidelines Research and Evaluation) instrument. Each guideline was assessed by two certified reviewers.
**Results**: Each reviewer spent 2-5 hours on each guideline. The average cost for evaluating one guideline amounted to about €800. Fifteen guidelines received the conclusion "strongly agree", 98 "recommend with provisos or alterations", 8 "would not recommend", and 6 received "unsure".
**Conclusions:** Most Norwegian clinical guidelines do not fulfil the quality criteria in the AGREE instrument. Better guidance for rating the overall assessment is needed. Systematic quality review of guidelines is more structured than peer review of scientific articles, but has less consequence as it is done independently of publication. Guidelines should be reviewed by an independent body before publication, and their evaluation should include novelty and relevance.

**Keywords** Practice guidelines as topic; quality control; guideline adherence; review, systematic; Norway; AGREE

### Background

The Institute of Medicine, part of the National Academy of Sciences in the United States of America, has defined clinical guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances".[1] The production and use of clinical guidelines are increasing. Sixty one per cent of general practitioners in the Netherlands[2] and 93% in Great Britain[3] reported using guidelines that they believed matched the needs of individual patients. Clinical guidelines are primarily intended to improve the quality of care,[4,5] but are also used as indicators of health-care quality[6,7] as well as for legal, political, and economic purposes. They are frequently published on open internet sites and in guideline databases without a systematic quality review prior to publication.

Studies within several medical fields have shown that the quality of clinical guidelines from central guideline developers in Europe, North America, and international organisations varies and that most do not fulfil high quality criteria.[8-13] There is a huge imbalance between the number of guidelines and the number of high-quality studies that assess their effectiveness and effect on patient outcomes.[5]

To assess the quality of guidelines, the AGREE (Appraisal of Guidelines Research and Evaluation) instrument was developed by an international collaboration of researchers and guideline developers funded by the BIOMED-2 programme of the European Union[14] and validated on 100 guidelines from 11 countries, with 195 appraisers. The Guidelines International Network (G-I-N), a global not-for-profit association which holds the world's largest international guideline library, works in partnership with the AGREE Research Trust.

Several questions regarding the evaluation of guidelines remain to be answered. Should guidelines be evaluated before or after they are published? Could and should the AGREE instrument[15] be used as a template for guideline development or as a tool for evaluating guidelines after publication? Will evaluations with AGREE help to identify the most reliable recommendations in the treatment of individual patients? Will any potential benefits of AGREE evaluations justify their cost? In addition, can we – in guideline publication and evaluation – learn something from the long-established tradition of using peer review to assess the quality of scientific articles?

The objectives of this study were to test the feasibility and the benefits of systematic quality reviews of clinical guidelines. We have also compared this process with that of peer review of scientific articles.

### Methods

The Norwegian Electronic Health Library (NEHL) includes a database with 457 Norwegian clinical guidelines, guidance documents, and procedures. We selected clinical guidelines for quality review, using the following inclusion criteria: Norwegian origin of guideline, publication in the period 2000–2009, comprehensive clinical guidelines (not procedures or summaries), national (not only local) relevance, and direct relation to patient care.

We used the AGREE instrument[15] for the evaluation. AGREE evaluations assess 23 items (see box), each of which

**Domains and items in the AGREE instrument**

*Scope and purpose*

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*

1 The guideline development group includes individuals from all the relevant professional groups.

2 The patients' views and preferences have been sought.

3 The target users of the guideline are clearly defined.

4 The guideline has been piloted among target users.

*Rigour of development*

1 Systematic methods were used to search for evidence.

2 The criteria for selecting the evidence are clearly described.

3 The methods used for formulating the recommendations are clearly described.

4 The health benefits, side effects and risks have been considered in formulating the recommendations.

5 There is an explicit link between the recommendations and the supporting evidence.

6 The guideline has been externally reviewed by experts prior to its publication.

7 A procedure for updating the guideline is provided.

*Clarity and presentation*

1 The recommendations are specific and unambiguous.

2 The different options for management of the condition are clearly presented.

3 Key recommendations are easily identifiable.

4 The guideline is supported with tools for application.

*Applicability*

1 The potential organizational barriers in applying the recommendations have been discussed.

2 The potential cost implications of applying the recommendations have been considered.

3 The guideline presents key review criteria for monitoring and/or audit purposes.

*Editorial independence*

1 The guideline is editorially independent from the funding body.

2 Conflicts of interest of guideline development members have been recorded.

is rated on a four point scale: 4 denotes "strongly agree", 3 "agree", 2 "disagree", and 1 "strongly disagree". The items are organized in six domains (box); "scope and purpose", "stakeholder involvement", "rigour of development", "clarity and presentation", "applicability", and "editorial independence". Each domain describes a dimension of the guideline quality. Domain scores were calculated by summing up all the scores of the individual items in a domain and by standardising the total as a percentage of the maximum possible score for that domain.

Nine doctors and one nurse took part in a one-day workshop followed by two training guideline assessments and were certified as AGREE reviewers. They undertook to review the selected guidelines, and each guideline was assessed by two reviewers during 2007–2009. The six domain scores were calculated with the aid of the formula: (obtained score – minimum possible score)/(maximum possible score – minimum possible score) × 100%. The guidelines received one out of four final conclusions: strongly recommend, recommend with alterations or provisos, would not recommend, or unsure. The two reviewers had to agree on the conclusion, for each assessment.

The results of the evaluations were published together with each guideline on our website, http://www.helsebiblioteket.no/Retningslinjer. Guideline developers were contacted before publishing the evaluation results and were given the opportunity to reply and to supply further information.

We used the AGREE instrument to evaluate the quality of Norwegian guidelines and to evaluate the feasibility and usefulness of doing these evaluations. Calculations of average scores with confidence intervals were performed using statistical software (SPSS 15).

## Results

Each reviewer spent from two to five hours on each clinical guideline they assessed. In total, 127 clinical guidelines were assessed. The average cost for evaluating one guideline amounted to 6400 Norwegian krone (€800), including administration costs.

The average scores (percentages; with standard error) for the six domains are shown in Table 1.

The average quality scores (mean (SE)) were high for "scope and purpose" and "clarity and presentation", which corresponds to "agree" or "strongly agree" on most of the items in those two domains. Those two domains also scored significantly higher than the mean score of 49% for all domains ($p<0.01$). "Rigour of development", "applicability", "editorial independence", and "stakeholder involvement" had low scores, with a score of "disagree" or "strongly disagree" for the majority of the items. When all domains were combined, 15 guidelines received the conclusion "strongly agree" (most scores above 60%), 98 guidelines received "recommend with provisos or alterations" (most scores between 30% and 60%), eight guidelines received

**Table 1: Average scores for the six domains**

| Domain | Mean (SE) (%) |
|---|---|
| Scope/purpose | 73 (3) |
| Clarity/ presentation | 60 (3) |
| Stakeholder involvement | 46 (3) |
| Editorial independence | 35 (3) |
| Applicability | 35 (3) |
| Rigour of development | 33 (3) |

"would not recommend", and the reviewers were "unsure" about six guidelines.

## Discussion

Most Norwegian guidelines did not fulfil all the quality criteria in the AGREE instrument: four out of six domains yielded low scores and three quarters received the overall assessment "recommend with provisos or alterations". It was cost and time consuming to do the evaluations.

We evaluated core clinical guidelines from a wide range of medical fields. The reviewers were trained and certified before assessing, but the AGREE instrument does not provide clear guidance on when to give 1, 2, 3, or 4 points on each item scored. Results of evaluations depend on how accurately authors have documented the work and material on which each guideline is based. Incomplete reporting and documentation may result in misleadingly low scores.

This project has provided Norwegian healthcare personnel with an overview of guidelines from different developers, together with the results of the quality evaluation. This has made the guideline quality more transparent, and we hope that this will help to raise the standard of existing as well as new guidelines. Our results support previous published studies, which show that the quality of guidelines varies and may be poor.[8-13] Few guidelines received a high score. For many conditions there are no Norwegian guidelines with high scores. Even with high scores, research has shown that different guidelines on the same topic can provide recommendations that differ considerably.[16,17] Factors other than scientific evidence – such as socioeconomics, cultural differences, or characteristics of health systems – can also influence the development of recommendations.[18] As one of the founders of the AGREE instrument points out: "the AGREE instrument does not assess the clinical content of the recommendations or the quality of the supporting evidence. Good methodological quality does not necessarily indicate good-quality recommendations."[19] We have confirmed the need for an appropriate review system that can help practitioners to find the most reliable guidelines. The feasibility of this project depends on the capacity of a third party to carry out the work. This involves administration, practical work, and costs incurred by the evaluation process, as well as publication of the results.

Table 2 reveals important differences between the process of systematic quality review of guidelines and traditional peer review of scientific manuscripts submitted to medical journals. The overview is based on instructions to authors and peer reviewers in major journals like *BMJ* and *The Lancet*[20,21] and guidelines for peer review from international organisations for medical editors.[22-25] Even if peer reviewers are not paid on a regular basis there are administration costs; analyses have estimated the cost of peer review to be $200-480,[26,27] but studies report that peer review costs vary substantially.[28]

In contrast to clinical guidelines, scientific articles are constructed according to an international standard, the IMRAD (introduction, methods, results, and discussion) structure,[29] and are usually peer reviewed according to guidelines on good publication practice.[16-19] But different journals use different review forms, which do not ensure that the same items are checked in each review. Reviewers can use the CONSORT (Consolidated Standards of Reporting Trials) statement, which is intended to improve the reporting of randomized controlled trials, and extensions

**Table 2. Quality review of clinical guidelines with AGREE (Appraisal of Guidelines Research and Evaluation) compared with peer review of scientific articles**

| Quality review of clinical guidelines | Peer review of scientific articles |
|---|---|
| Aiming at increased quality | Aiming at increased quality |
| Not routinely and systematically done | Routinely done for all scientific journals |
| Post-publication review | Pre-publication review |
| Reviewers receive systematic training and are certified. | Experienced researchers or clinicians do the reviewing without any formal training. |
| Universal review instrument available | Semi-structured review form for each journal |
| Reviewers are paid | Reviewers not paid on a regular basis |
| Reviewers know each other's identity and aim at agreeing on the final conclusion | Reviewers do not know each other and do not have to agree |
| Reviewers are not necessarily specialists in the medical field discussed in the guideline | Reviewers are normally specialists in the field of the paper |
| Review limited to purpose, stakeholder involvement, rigour of development, presentation, application and editorial independence | Review includes the whole research process, (hypotheses, design, data collection and analysis) as well as presentation and conclusion |
| The structure of clinical guidelines varies and makes evaluation difficult | Scientific articles usually follow the IMRAD structure which simplifies peer review |
| Change to clinical guideline based on quality review is voluntary and up to the guideline developers | Scientific articles are not published if reviewers and the editor of the journal are not satisfied |

of the CONSORT statement have also been developed for other types of study design.[30] Tom Jefferson, an author and reviewer in the Cochrane Collaboration, found that no methods for quality improvement other than peer review had been tested and suggested that this should be done.[27] The structured and transparent method of quality review of clinical guidelines with AGREE ensures that all items in the form will be checked in the course of each review process. Evaluation of guidelines with AGREE rests upon the evaluation of the methods and process alone; relevance, importance of the questions addressed in the guidelines, novelty, and validity of the medical content are not directly evaluated. This is a weakness compared with traditional peer review of scientific articles, where both the content and the process are evaluated, and it builds on an assumption that a good process leads to correct medical content.

Evaluation of guidelines with AGREE usually takes the form of a review of already-published guidelines, and these evaluations have little impact on the quality of the guidelines even though some developers – for example, the National Institute for Health and Clinical Excellence (NICE) – use AGREE as a checklist when they develop guidelines. National Clearinghouse, a public guideline library in the United States and part of the Agency for Healthcare Research and Quality,[32] uses some quality criteria for guidelines to be included in the database. The peer review process of scientific articles is done prior to publication and can therefore affect the quality of articles before they are issued and indeed determine whether they are published or not. A new version of AGREE has now been published with minor changes,[33] but its guidance on rating is no better and it does not include questions about novelty or relevance.

## Conclusions

Because the content, purpose, and structure of clinical guidelines vary, and differ from those of scientific articles, we need methods for appraisal that differ from those designed to evaluate scientific research.

Both processes, peer review of scientific articles and quality evaluation of guidelines, can be improved. Peer review of scientific articles can become more structured, transparent, and consistent. Guidelines should be reviewed by an independent body before publication, and the evaluation would benefit from including perspectives that relate to patient outcome, novelty, and relevance. From our findings we conclude that systematic quality review of clinical guidelines is feasible, and that the scores can be useful for guideline developers when updating guidelines, but the overall assessment "recommend with provisos or alterations" given to three out of four guidelines in our study is not very helpful for clinicians. Better guidance for rating the overall assessment is needed.

**Acknowledgements** We thank Trond H Bjørnerud for contributing to the acquisition of data, guidance of the reviewers, and development of the National Norwegian guideline database. We also thank the reviewers of the clinical guidelines: Hege Lorentzen, Torolf Vågan, Nicola Herzig, Bjørn Sturød, Solrun Holm, Jahanzeb Mughal, Kashif Faiz, Mark Fagen, Ingrid Marie Hardang, and Arne Jan Hjemsæter.

**Competing interests** None declared. The study was funded by The Norwegian Electronic Health Library at The Norwegian Knowledge Centre for Health Services.

**Authors' contributions** AHR and MN designed the study, contributed substantially to the acquisition of data, and analyzed and interpreted the findings. AHR drafted the manuscript, and MN revised the manuscript critically for important intellectual content. EM contributed to analysis, interpretation of data, and was involved in revising the manuscript critically for important intellectual content. All authors have read and approved the final manuscript.

## References

1 Committee to Advise the Public Health Service on Clinical Practice Guidelines, Institute of Medicine. *Clinical practice guidelines: directions for a new program*. Washington, DC: National Academy Press, 1990.

2 Grol R, Dalhuijsen J, Siep T, Veld C, Rutten G, Mokkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998;31(7162):858–861.

3 Watkins C, Harvey I, Langley C, Gray S, Faulkner A. General practitioners' use of guidelines in the consultation and their attitudes to them. *British Journal of General Practice* 1999;49(438):11–15.

4 Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *International Journal of Quality in Health Care* 2003;15(1):31–45.

5 Lugtenberg M, Burgers JS, Westert GP. Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. *Quality and Safety in Health Care* 2009;18(5):385–392.

6 Cluzeau F, Burgers J, Brouwers M, Grol R, Mäkelä M, Littlejohns P, et al. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality and Safety in Health Care* 2003;12(1):18–23.

7 Hermens RP, Ouwens MM, Vonk-Okhuijsen SY, van der Wel Y, Tjan-Heijnen VC, van den Broek LD, et al. Development of quality indicators for diagnosis and treatment of patients with non-small cell lung cancer: a first step toward implementing a multidisciplinary, evidence-based guideline. *Lung Cancer* 2006;54(1):117–124.

8 Peterson A, Carlhed R, Lindahl B, Lindstrom G, Aberg C, Andersson-Gare B, et al. Improving guideline adherence through intensive quality improvement and the use of a National Quality Register in Sweden for acute myocardial infarction. *Quality Management in Health Care* 2007;16(1):25–37.

9 Nagy E, Watine J, Bunting PS, Onody R, Oosterhuis WP, Rogic D, et al. Do guidelines for the diagnosis and monitoring of diabetes mellitus fulfill the criteria of evidence-based guideline development? *Clinical Chemistry* 2008;54(11):1872–1882.

10 Navarro Puerto MA, Ibarluzea IG, Ruiz OG, Alvarez FM, Herreros RG, Pintiado RE et al. Analysis of the quality of

clinical practice guidelines on established ischemic stroke. *International Journal of Technology Assessment in Health Care* 2008;24(3):333–341.
11 Nelen WL, van der Pluijm RW, Hermens RP, Bergh C, de Sutter P. Nygren KG, et al. The methodological quality of clinical guidelines of the European Society of Human Reproduction and Embryology (ESHRE). *Human Reproduction* 2008;23(8):1786–1792.
12 Lopez-Olivo MA, Kallen MA, Ortiz Z, Skidmore B, Suarez-Almazor ME. Quality appraisal of clinical practice guidelines and consensus statements on the use of biologic agents in rheumatoid arthritis: a systematic review. *Arthritis and Rheumatism* 2008;15(11):1625–1638.
13 Misso ML, Pitt VJ, Jones KM, Barnes HN, Piterman L, Green SE. Quality and consistency of clinical practice guidelines for diagnosis and management of osteoarthritis of the hip and knee: a descriptive overview of published guidelines. *Medical Journal of Australia* 2008;189(7):394–399.
14 Vitry AI, Zhang Y. Quality of Australian clinical guidelines and relevance to the care of older people with multiple comorbid conditions. *Medical Journal of Australia* 2008;189(7):360–365.
15 AGREE Collaboration. Appraisal of Guidelines for Research and Evaluation Instrument (AGREE). 1991. http://www.agreecollaboration.org/pdf/aitraining.pdf. Appendix 1. (Accessed 13 January 2011.)
16 Haagen EC, Hermens RP, Nelen WL, Braat DD, Grol RP, Kremer JA. Subfertility guidelines in Europe: the quantity and quality of intrauterine insemination guidelines. *Human Reproduction* 2006:21(8):2103–2109.
17 McIntosh HM, Neal RD, Rose N, Watson E, Wilkinson C, Weller D, et al. Follow-up care for men with prostate cancer and the role of primary care: a systematic review of international guidelines *British Journal of Cancer* 2009;100(12):1852–1860.
18 Burgers JS, Bailey JV, Klazinga NS, van der Bij AK, Grol R, Feder G. Inside guidelines: comparative analysis of recommendations and evidence in diabetic guidelines from 13 countries. *Diabetes Care* 2002;25(11):1933–1939.
19 Burgers JS. Guideline quality and guideline content: are they related. *Clinical Chemistry* 2006;52(1):3–4.
20 TheLancet.com. For authors. http://www.thelancet.com/for-authors. (Accessed 13 January 2011.)
21 BMJ. Guidance for peer reviewers. http://resources.bmj.com/bmj/reviewers/peer-reviewers-guidance. (Accessed 12 Jan 2011.)
22 Equator Network. Resources for editors and peer reviewers. 30 June 2010. http://www.equator-network.org/resource-centre/editors-and-peer-reviewers/editors-and-peer-reviewers/#editperr. (Accessed 13 January 2011.)
23 International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publications. http://www.icmje.org. (Accessed 13 January 2011.)
24 World Association of Medical Editors. Policy statements. http://www.wame.org/resources/policies. (Accessed 13 January 2011.)
25 Committee on Publication Ethics. Code of conduct. http://publicationethics.org/code-conduct. (Accessed 13 January 2011.)
26 Rowland F. The peer-review process. *Learned Publishing* 2002;15(4):247–258.
27 King DW, Tenopir C. Scholarly journal and digital database pricing: threat or opportunity? Paper presented at PEAK 2000 conference, Michigan, 23–24 March 2000.
28 King DW. The cost of journal publishing: a literature review and commentary. *Learned Publishing* 2007;20:85–106.
29 Sollaci LB, Peira MG. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *Journal of the Medical Libraries Association* 2004;92(3):364–371.
30 CONSORT. The CONSORT Statement. 10 November 2010. http://www.consort-statement.org/consort-statement. (Accessed 13 January 2011.)
31 Jefferson T. Quality and value: models of quality control for scientific research. A look at alternatives to the peer review system suggests it might be the best bet. *Nature* 2006; doi:10.1038/05031 (http://www.nature.com/nature/peerreview/debate/nature05031.html).
32 Agency for Healthcare Research and Quality. National Guidelines Clearinghouse. http://www.guideline.gov/index.aspx.
33 AGREE II. 2010. Downloadable at http://www.agreetrust.org.

## Scientific discourse and contrastive linguistics: hedging

**Françoise Salager-Meyer**
*Graduate School of Medicine, University of The Andes, Mérida, Venezuela; francoise.sm@gmail.com*

**Abstract** Expressing tentativeness and possibility – presenting claims with caution and precision – is important in research papers. Writers from different linguistic backgrounds show variations in the amount of this "hedging" and the types of devices they use.

**Keywords** Science; linguistic hedges; claim; modalisation; knowledge; politeness; mitigation

Hedging is the expression of tentativeness and possibility. It is therefore central to academic/scientific writing, where statements are rarely made without subjective assessment of their reliability and where claims need to be presented with caution and precision. Science indeed is scepticism, doubt, refutation, speculation, formulation of hypothesis, criticism. As a consequence, the expression of doubt and possibility is central to the negotiation of claims, and what counts as effective persuasion is influenced by the fact that evidence, observations, data, and flashes of insight must be shaped with due regard for the nature of reality and their acceptability to an audience.

In medical writing, hedges play a critical role in gaining ratification for claims for a powerful peer group by allowing writers to present statements with appropriate accuracy,