

Evidence-Based Practice Guidelines in OHS: Are They Agree-Able?^a

Carel HULSHOF^{1,2*} and John HOENEN³

¹ Netherlands Society of Occupational Medicine (NVAB), Centre of Excellence, PO Box 2113, 3500 GC Utrecht, The Netherlands

² Academic Medical Center, Universiteit van Amsterdam, Coronel Institute of Occupational Health, PO Box 22700, 1100 DE Amsterdam, The Netherlands

³ Bureau Beter Consultancy, Potgieterstraat 34, 3532 VS Utrecht, The Netherlands

Received August 4, 2006 and accepted November 2, 2006

Abstract: The purpose of this study was to evaluate the acceptance, validity, reliability and feasibility of the AGREE (Appraisal of Guidelines and REsearch and Evaluation) instrument to assess the quality of evidence-based practice guidelines for occupational physicians. In total, 6 practice guidelines of the Netherlands Society of Occupational Medicine (NVAB) were appraised by 20 occupational health professionals and experts in guideline development or implementation. Although appraisers often disagreed on individual item scores, the internal consistency and interrater reliability for most domains was sufficient. The AGREE criteria were in general considered relevant and no major suggestions for additional items for use in the context of occupational health were brought up. The domain scores for the individual guidelines show a wide variety: ‘applicability’ had on average the lowest mean score (53%) while ‘scope and purpose’ had the highest one (87%). Low scores indicate where improvements are possible and necessary, e.g. by providing more information about the development. Key experts in occupational health report that AGREE is a relevant and easy to use instrument to evaluate quality aspects and the included criteria provide a good framework to develop or update evidence-based practice guidelines in the field of occupational health.

Key words: Practice guideline, OHS, Quality, AGREE, Evidence-based medicine

Introduction

In the Netherlands, the Dutch government passed legislation in 1995 that required all employers to contract certified multidisciplinary occupational health services (OHS) to assist them in occupational health and safety and in sickness absence management¹. The coverage of the working population by OHS increased within a few years from 40% to 95%. More than 2000 occupational physicians (OPs) and other professionals deliver occupational health care for the working population of about 7 million people. Occupational health care for all workers can be considered

as a substantial progress. However, the dominating focus within the work of Dutch OHS is on sickness absence management and less on prevention. The quality of the care provided by the OHS often does not meet professional standards, partly as a result of the commercial approach that many services adopt to survive in a rapidly changing market. Moreover, the direct influence of ‘third parties’ (e.g. insurance companies or commercial providers of certain aspects of care) on the content of the work of OPs and other professionals is growing. Therefore, consensus exists about the need for enhancement of the scientific basis of occupational health practice in order to counteract this development and to improve the professional quality and professional independence of professionals within the OHS². One of the tools for this is the development, implementation,

^aPresented at the 13th International Congress on Occupational Health Services, November 30 – December 3, 2005, Utsunomiya, Japan.

*To whom correspondence should be addressed.

and evaluation of clinical practice guidelines. Practice guidelines are 'systematically developed statements designed to assist practitioner decisions about appropriate health care for specific clinical circumstances'³). Such guidelines are particularly useful if there is a large variation in current practice; if they contain new evidence with an important impact on health management; or if they affect many individuals at high risk or involve such high costs that even small changes in practice could have major impact on health outcomes or resources⁴). In 1998, the Netherlands Society of Occupational Medicine (NVAB) started a programme for the development and implementation of evidence-based practice guidelines for occupational physicians. The NVAB-guidelines are based on scientific evidence (for each important recommendation in the guideline the level of evidence is presented), consensus in a peer group, professional or ethical principles, and best practices. So far, seven NVAB-guidelines (on low back pain, mental health, visual acuity, sheltered workshops, complaints of arm, neck or shoulders, asthma/COPD, and contact dermatitis) have been published and five others are currently in development. Besides this, the NVAB is also actively involved in developing multidisciplinary guidelines in collaboration with other occupational health professionals, with general practitioners, and with other medical specialists^{5, 6}). Guidelines are not only of use for the individual practitioner, they may improve the quality of health care by enhancing professionalization, accountability and efficiency⁷). Developing guidelines is laborious and expensive. Each guideline development process lasts in general one and a half year. The enormous proliferation of guidelines in health care and concerns about their quality has led to the international development of the Appraisal of Guidelines and Research and Evaluation (AGREE) instrument, a validated generic instrument for the appraisal of guidelines⁸). It is officially recommended by the Council of Europe and by the WHO to help both guideline developers and users to assess the methodological quality of clinical practice guidelines. This instrument was mainly developed and tested in guidelines in primary and secondary clinical health care.

The field of occupational health is considered lagging behind other medical specialties in failing to incorporate evidence-based medicine into practice⁹). The setting and content of occupational health care is, however, different from clinical care or general practice. The work of OP's is often strongly influenced by its contextual framework, in particular by national legislation on work and health¹⁰). Much practice derives from routine procedures or legal requirements. Therefore, in OPs a compliance mentality is

frequent where action is often taken without an urge to explore its basis¹¹). Clinical questions in occupational health practice are often not dealing with diseases but with risks, impairments, disabilities, return to work or lifestyle. For some of these issues, the developed clinical methods for evidence-based search strategies of the literature are insufficient and additional strategies for access to international and national databases have to be elaborated^{12, 13}). Another impediment to the application of evidence-based medicine (EBM), and consequently also for the development of practice guidelines, is the lack of research evidence in the field. In particular, robust evidence on the efficacy or effectiveness of occupational health activities performed by OP's is limited^{2, 14}).

On the other hand, we should not be too pessimistic. Recent initiatives in the application of EBM in daily occupational health practice are promising and show that there is more field-relevant evidence available in e.g. Medline than is generally assumed by OPs¹⁵⁻¹⁸). Moreover, a cohort study and two controlled trials with occupational health guidelines on low back pain and mental health problems confirmed their favourable effect on sick leave and prevention of disability¹⁹⁻²¹).

Therefore, in order to explore if the process of guideline development for OP's is still 'lagging behind', we evaluated the validity, reliability, feasibility, and acceptance of the AGREE instrument to assess the quality of the clinical practice guidelines published by the NVAB. This evaluation was part of a larger project which had a broader focus: to study if the AGREE instrument, specifically developed and validated to appraise the quality of clinical practice guidelines, is feasible and useful for the evaluation of a diversity of 'knowledge documents' for occupational health professionals, also for documents that were not explicitly developed as a clinical practice guideline for physicians. The results of this larger project are published elsewhere²²). In this paper, we, however, specifically focus on the usability of the AGREE instrument for evaluating the quality of clinical practice guidelines for OP's: the NVAB-practice guidelines.

Material and Methods

Selection of guidelines and appraisers

Covering a broad range of occupational health issues, in total 22 knowledge documents from different organizations in occupational health were selected. Inclusion criteria were: published during last 5 years, indications for regular use in daily practice or continuous medical examination, dealing with prevalent occupational health issues, targeted to occupational health professionals, and publicly availability of the document. From the 22 selected documents, 6 were

NVAB practice guidelines. We made a selection of 4 monodisciplinary guidelines, 1 collaborative guideline and 1 more general guideline. These guidelines were dealing with respectively low back pain; mental health disorders; asthma and COPD; complaints of arm, neck, and shoulders; collaboration with general practitioners in the case of workers with mental health disorders; and referrals to medical specialists.

Per document, at least four independent appraisers (occupational health professionals and experts in guideline development, evidence-based medicine, or implementation) with many years of experience in the field of occupational health were recruited. In total, 31 appraisers volunteered in the project; 20 of them appraised one or more NVAB practice guideline. This last group of appraisers consisted of 10 occupational physicians, 4 researchers in the field of occupational health, and 6 other professionals like occupational hygienist, occupational psychologist, insurance physician, or educational expert. Each of the NVAB practice guidelines was on average appraised by 4 different volunteers. Due to non response, the practice guideline on mental health disorders was appraised by only 2 appraisers. Most of the appraisers judged three different documents. To avoid conflict of interest, members of the guideline development group, staff members of the NVAB Centre of Excellence and external referees of the guidelines were not eligible to appraise their 'own' guidelines.

The instrument: AGREE

The AGREE instrument exists of 24 items, 23 of them grouped into 6 quality domains: *scope and purpose* (the overall aim, the specific clinical questions and the target population); *stakeholder involvement* (the extent to which the guideline represents the view of its intended users); *rigor of development* (the methodology of reviewing and synthesizing the evidence, and formulating and updating the recommendations); *clarity and presentation* (language and format); *applicability* (the likely organizational, behavioural, and cost implications); and *editorial independence* (independence of recommendations and acknowledgement of possible conflict of interest from the guideline development group)²³. The last question asks if the appraiser overall recommends the guideline for use in daily practice. Each appraiser scores the extent to which every item (criterion) had been fulfilled on a 4-point Likert scale from 'strongly agree' to 'strongly disagree'. Domain scores are calculated by summing up all the scores of the individual items in a domain and by standardizing the total as a percentage of the maximum possible score for that domain:

$$\text{standardized domain score} = \frac{\text{obtained score} - \text{minimum possible score}}{\text{maximum possible score} - \text{minimum possible score}} \times 100$$

In addition to the AGREE score, the appraisers were also asked to rate on a 4-point Likert scale added questions about the relevance, the usability and the appreciation of the different components of the instrument for use in the field of occupational health.

Data analysis

Domain scores were calculated according to the instructions of the AGREE instrument, in which 0% (all items scored 'strongly disagree' by all appraisers) is the minimum score and 100% (all strongly agree) the maximum score. The internal consistency of each domain was evaluated by using Cronbach's alpha and the reliability between appraisers by calculating intraclass correlation coefficients (ICC). The significance of differences in mean domain scores was determined by one-way analysis of variance and differences between subgroups on item level by non-parametric Kruskal-Wallis H test. All analyses were performed, using SPSS 10.0.7.

Results

The domain scores for the individual NVAB practice guidelines show a wide variety, ranging from 36.1 to 100% (Table 1). On average, over the selected practice guidelines, the domain 'applicability' had the lowest mean score (53%) while 'scope and purpose' had the highest one (87%). Some items scored low for most of the guidelines, e.g. 'the patients' views and preferences have been sought' (on average 2.5 on the scale between 1 and 4); 'a procedure for updating the guideline is provided' (2.1), and 'conflicts of interest of guideline development members have been recorded' (2.3).

Internal consistency for most domain scores was good: Cronbach's alpha per domain score for our total study (not limited to the NVAB practice guidelines) ranged from 0.62 to 0.87. Appraisers often disagreed on item scores. The interrater reliability was moderate: ICC mean for domain scores by 4 appraisers ranged from 0.49 to 0.81. (Table 2).

The relevance of the AGREE-items for assessing the quality of the NVAB practice guidelines was rated high: between 3.1 and 3.9 on a scale from 1 to 4. The relevance scores were independent of the assessed guideline or the background of the appraiser.

Most of the appraisers of the NVAB guidelines thought that the AGREE-instrument was easy to use: on average

Table 1. AGREE domain scores (in%) of the appraised NVAB practice guidelines

Domain	LBP	LESA	CANS	Mental health	Asthma/ COPD	Referrals
Scope and purpose	94.4	74.1	86.1	100	91.7	86.1
Stakeholder involvement	58.3	58.3	60.4	70.8	72.9	81.3
Rigor of development	40.5	43.7	69.1	59.5	78.6	58.3
Clarity and presentation	89.6	56.9	77.1	62.5	77.1	93.8
Applicability	44.4	38.9	36.1	66.7	61.1	86.1
Editorial independence	50.0	66.7	66.7	50.0	50.0	62.5
<i>Number of appraisers</i>	4	6	4	2	4	4

LBP: Low Back Pain. LESA: joint guideline with Dutch College of General Practitioners (NHG) on treatment of adjustment disorders and burnout. CANS: Complaints of Arm, Neck, and Shoulders. Referrals refers to a protocol on referrals for Occupational Physicians.

Table 2. Internal consistency and interrater reliability for each domain

	Cronbach α	ICC (4 raters)
Scope and Purpose (3 items)	0.85	0.70
Stakeholder involvement (4 items)	0.65	0.62
Rigor of development (7 items)	0.87	0.81
Clarity and presentation (4 items)	0.75	0.62
Applicability (3 items)	0.76	0.72
Editorial independence (2 items)	0.62	0.49
<i>Number of appraised guidelines</i>	6	5*

*One guideline was appraised by only 2 raters.

3.4 (range 2.83–3.67). The appreciation for AGREE was in general positive: 3.2 (2.91–3.69). Only few suggestions for additional items (e.g. on collaboration between occupational and curative health care) for use in the specific context of occupational health were brought up.

Discussion

Key persons in the field of occupational health agree on the importance of the AGREE criteria for assessing the quality of practice guidelines for OP's in the Netherlands. The findings suggest that the AGREE instrument is also reliable and valid for this purpose. Both internal consistency and interrater reliability are well in line with the original validation study by the AGREE Collaboration⁸. This is in agreement with other recent studies that have tested AGREE in medical or paramedical fields of health care where the original instrument was not developed^{24–26}.

It should be kept in mind that the instrument evaluates process aspects of the quality of a guideline and not the content but also measures how well the guideline

development process is reported. For some of the items, e.g. 'a procedure for updating the guideline is provided', a negative score will not per se mean that such a procedure is not available but only that it is not reported in the guideline. Nevertheless, low item and domain scores for the individual practice guidelines indicate where improvements are possible and necessary. The results of our study indicate that there is ample room for improving the quality of the NVAB guidelines. Comparison of the mean domain scores of the NVAB guidelines with the ones from experienced clinical practice guideline organisations^{27, 28} show that the NVAB scores tend to be higher (Table 3). Although such a comparison has many limitations, this is at least encouraging.

In the field of occupational health, Cates *et al.*²⁵ recently evaluated the American ACOEM Guidelines with the use of AGREE and concluded that future ACOEM guidelines should incorporate better reporting and give closer attention to the guideline process. Figure 1 shows the comparison between the domain scores between the American and Dutch occupational health practice guidelines. In the American study, a very high inter-examiner agreement on item level for the (medical and chiropractic) assessors was found. This is in contrast with our findings but it may be due to the fact that the appraisers in that study all had prior guideline evaluation experience and were already familiar with the AGREE instrument.

In conclusion, AGREE is a relevant and easy to use instrument to assess quality aspects of occupational practice guidelines. The 6 domains cover a broad range of aspects of quality. Increasing the number of appraisers from 2 to 4 leads to an increase in internal consistency and interrater agreement²². This is in agreement with other studies^{8, 24, 27} For reliable scores, we, therefore, recommend to use 4 appraisers. Proper training of the appraisers in using the

Table 3. Mean AGREE domain scores of NVAB guidelines versus other clinical practice guidelines

Domain	AGREE*	Oncology**	NHG***	CBO***	NVAB
Scope and purpose	69.3	63.2	57.8	53.3	86.6
Stakeholder involvement	36.1	34.3	53.3	61.7	66.0
Rigor of development	40.7	42.2	33.3	60.0	56.9
Clarity and presentation	65.8	57.2	80.0	68.3	75.7
Applicability	36.9	25.8	13.3	44.4	52.7
Editorial independence	30.3	47.5	0.00	43.3	54.9
<i>Number of guidelines</i>	33	32	5	5	6

*Original validation study of the AGREE instrument⁸⁾. **International assessment of quality of oncology guidelines²⁷⁾. ***Assessment of quality of NHG and CBO guidelines²⁸⁾.

NHG: The Dutch Society of General Practitioners. CBO: The Dutch Institute for Healthcare Improvement.

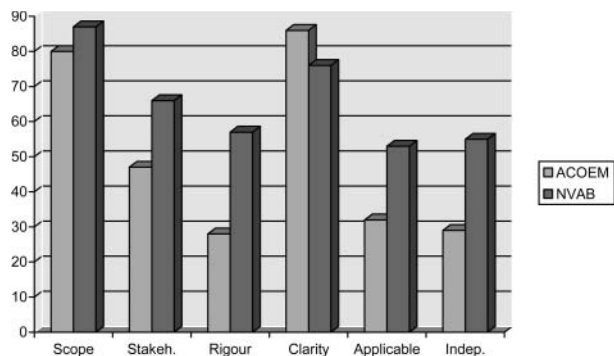


Fig. 1. Comparison of the mean AGREE domain scores of the American College of Occupational and Environmental Medicine (ACOEM) Guidelines as reported by Cates *et al.*²⁶⁾ and the Practice Guidelines of the Netherlands Society of Occupational Medicine (NVAB) as reported in this study.

AGREE could improve reliability. Like most appraisers in our study, we think the AGREE criteria could also be used by guideline developers as a framework to develop or update practice guidelines and in doing so improve their quality substantially.

Acknowledgements

This work was supported by a grant from the Dutch Ministry of Social Affairs and Employment. We have to thank all appraisers who shared their valuable time with us.

Conflict of Interest

Carel Hulshof is as co-ordinator practice guidelines of the NVAB highly involved in the development of the NVAB guidelines; John Hoenen was from 1999–2002 managing

supervisor of the ‘Guideline Bureau’ (now: Centre of Excellence) of the NVAB. Both authors were, however, not involved as appraisers in this study.

References

- 1) Weel AN, van der Beek AJ, Kroon PJ, Verbeek JH, van Dijk FJ (1999) Recent changes in occupational medicine in The Netherlands. *Int Arch Occup Environ Health* **72**, 285–91.
- 2) Hulshof CTJ, Verbeek JHAM, van Dijk FJH, Braam ITJ, van der Weide WE (1999) Evaluation research in occupational health services: general principles and a systematic review of empirical studies. *Occup Environ Med* **56**, 361–77.
- 3) Grimshaw JM, Russell IT (1993) Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* **342**, 1317–22.
- 4) Sackett DL, Strauss SE, Richardson WS (2000) Evidence-based medicine. How to practice and teach EBM, 2nd Ed., Churchill Livingstone, Edinburgh.
- 5) Romeijnders ACM, Vriezen JA, van der Klink JJJ, Hulshof CTJ, Terluin B, Flikweert S, Baart PC (2005) National primary care NHG-NVAB collaboration guideline ‘adjustment disorder’. *Huisarts Wet* **48**, 20–3 (in Dutch).
- 6) Verbeek J, Valk M, Kramer E, Lebbink M, Robeer GG, van Rooy GBGJ, Wiers HJ, Fortuin RJ, van Barneveld TA (2004) Structural attention for work in the development of multidisciplinary CBO-guidelines. *Tijdschr Bedrijfs Verzekeringsgeneesk* **12**, 343–5 (in Dutch).
- 7) Klazinga N (1994) Compliance with practice guidelines: clinical autonomy revisited. *Health Policy* **28**, 51–66.
- 8) Agree Collaboration: Writing Group (Cluzeau F, Burgers J, Brouwers M, *et al.*) (2003) Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines. *Qual Saf Health Care* **12**, 18–23.
- 9) Birrell L, Beach J (2001) Developing evidence-based guidelines in occupational health. *Occup Med* **51**, 73–4.
- 10) Van Dijk FJH (2000) Aims and contents of modern

- occupational health services. *People and Work Research Report* **38**, 27–35.
- 11) Carter T (2000) The application of the methods of evidence-based practice to occupational health. *Occup Med* **50**, 231–6.
 - 12) Verbeek J, Salmi J, Pasternack I, Jauhiainen M, Laamanen I, Schaafsma F, Hulshof C, van Dijk F (2005) A search strategy for occupational health intervention studies. *Occup Environ Med* **62**, 682–7.
 - 13) Schaafsma F, Hulshof C, Verbeek J, Bos J, Dyserinch H, van Dijk F (2006) Developing search strategies in Medline on the occupational origin of diseases. *Amer J Ind Med* **49**, 127–37.
 - 14) Pransky G, Himmelstein J (1996) Outcomes research: implications for occupational health. *Am J Ind Med* **29**, 573–83.
 - 15) Verbeek JH, van Dijk FJ, Malmivaara A, Hulshof CT, Räsänen K, Kankaanpää EE, Mukala K (2002) Evidence-based medicine for occupational health. *Scand J Work Environ Health* **28**, 197–204.
 - 16) Grandi P, Franco G (2005) Practising evidence-based occupational health in workers' groups: how to prevent sickness absence caused by influenza. *Occup Med* **55**, 7–9.
 - 17) Verbeek J, van Dijk F (Eds.) (2006) A practical guide for the use of research information to improve the quality of occupational health practice. In: *Protecting Workers' Health Series No. 7*, WHO, Geneva.
 - 18) Schonstein E, Verbeek JH (2006) Occupational health systematic reviews: an overview. *Work* **26**, 255–8.
 - 19) Van der Weide WE, Verbeek JHAM, Van Dijk FJH (1999) Relation between indicators for quality of occupational rehabilitation of employees with low back pain. *Occup Environ Med* **56**, 488–93.
 - 20) Nieuwenhuijsen K, Verbeek JHAM, Siemerink J, Tummers-Nijssen D (2003) Quality of rehabilitation among workers with adjustment disorders according to practice guidelines; a retrospective cohort study. *Occup Environ Med* **60** (Suppl 1), i21–5.
 - 21) Van der Klink JJ, Blonk RW, Schene AH, van Dijk FJ (2003) Reducing long term sickness absence by an activating intervention in adjustment disorders: a cluster randomised controlled design. *Occup Environ Med* **60**, 429–37.
 - 22) Hoenen JAHJ, Hulshof CTJ, Burgers JS (2007) Appraisal of the quality of knowledge documents in occupational health. An assessment of the usefulness of the AGREE instrument. *Tijdschr Bedrijfs- en Verzekeringsgeneesk* (accepted for publication) (in Dutch).
 - 23) The AGREE Collaboration. Appraisal of Guidelines for Research & Evaluation (AGREE) instrument. <http://www.agreecollaboration.org>. Accessed August 1, 2003.
 - 24) MacDermid JC, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham ID (2005) Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. *BMC Health services Research* **5**, 18.
 - 25) Cates JR, Young DN, Guerriero DJ, Jahn WT, Armine JP, Korbett AB, Bowerman DS, Porter RC, Sandman T, King RA (2003) An independent assessment of chiropractic practice guidelines. *J Manipul Physiol Ther* **26**, 282–6.
 - 26) Cates JR, Young DN, Bowerman DS, Porter RC (2006) An independent AGREE evaluation of the Occupational Medicine Practice Guidelines. *The Spine Journal* **6**, 72–7.
 - 27) Burgers JS, Fervers B, Haugh M, Brouwers M, Browman G, Philip T, Cluzeau FA (2004) International assessment of the quality of clinical practice guidelines in oncology using the Appraisal of Guidelines and Research and Evaluation instrument. *J Clin Oncol* **22**, 2000–7.
 - 28) Burgers JS, Boluyt N (2004) An assessment using the AGREE instrument of the quality of clinical practice guidelines produced by the Dutch Institute for Healthcare Improvement (CBO) and the Dutch College of General Practitioners (NHG) evaluation. *Huisarts Wet* **47**, 394–9 (in Dutch).