Review

The Royal Society of Medicine

Journal of the Royal Society of Medicine; 106(8) 315–322
DOI: 10.1177/0141076813486261

Guidelines in cardiac clinical practice: evaluation of their methodological quality using the AGREE II instrument

Sanjeeve Sabharwal[1], Vanash Patel[1], Sukhjinder S Nijjer[2], Ali Kirresh[1], Ara Darzi[1], John C Chambers[3], Iqbal Malik[4], Jaspal S Kooner[2] and Thanos Athanasiou[1]

[1]Department of Surgery and Cancer, Imperial College, London W2 1NY, UK
[2]National Heart and Lung Institute, Imperial College, London W2 1NY, UK
[3]School of Public Health, Imperial College, London W2 1NY, UK
[4]Cardiovascular Sciences and the Renal Institute, Imperial College, London W2 1NY, UK
Corresponding author: Sanjeeve Sabharwal. Email: sanjeeve.sabharwal@ic.ac.uk

Summary

Although clinical guidelines have an influential role in healthcare practice, their development process and the evidence they cite has been subject to criticism. This study evaluates the quality of guidelines in cardiac clinical practice by examining how they adhere to validated methodological standards in guideline development. A structured review of cardiac clinical practice guidelines published in seven cardiovascular journals between January 2001 and May 2011 was performed. The AGREE II assessment tool was used by two researchers to evaluate guideline quality. A total of 101 guidelines were identified. Assessment of guidelines using AGREE II found methodological quality to be highly variable (median score, 58.70%; range, 45.34–76.40%). 'Scope and purpose' (median score, 86.1%) and 'clarity of development' (median score, 83.3 %) were the two domains within AGREE II that received the highest scores. Applicability (median score, 20.80%; range, 4.20–54.20%) and editorial independence (median score, 33.30%; range, 0–62.50%) had the lowest scores. There is considerable variability in the quality of cardiac clinical practice guidelines and this has not improved over the last 10 years. Incorporating validated guideline assessment tools, such as AGREE II, may improve the quality of guidelines.

Keywords

AGREE II, cardiac practice, guidelines

Introduction

Clinical guidelines have been defined by the Institute of Medicine as 'statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options'.[1] Although evidence suggests clinical guidelines improve quality of clinical practice,[2] questions remain over the quality of guidelines and guideline development.[3,4]

It is recognized that published guidelines should be based on the best available evidence.[5,6] Randomized controlled trials provide a high level of evidence and are favourably viewed as a basis for guideline development.[7] Many guidelines, however, fail to cite randomized controlled trials and appear to be developed from lower levels of evidence or expert opinion.[6,8] Although this may occur because of a lack of available evidence within a certain field, concern on guideline quality can arise because guideline committees may be subject to bias which affects the recommendations they make.[9] Furthermore, recommendations on a single subject by different organizations have also been variable.[10] These factors highlight the need for rigour and uniformity in guideline development and presentation. Despite recommendations on appropriate methodology for guideline development,[11,12] a previous study evaluating guideline quality between 1985 and 1997, showed guidelines do not adhere well to established methodological standards.[13]

A further area in guideline development of considerable importance is whether conflicts of interest are reported.[14] Guidelines have a considerable impact on healthcare, and reporting on potential conflicts of interests would be a step towards improved transparency in their development process. However, a study of 191 guidelines found that such transparency is commonly neglected.[14] Recent consensus on guideline development focused upon the need to record conflicts of interest of members of the guideline development group, but failed to suggest these details be published within the guideline.[15] Reporting on conflicts of interests is also a significant component within the AGREE II assessment tool, a validated questionnaire that is used to assess the methodological quality of clinical practice guidelines (CPG).[16,17] A systematic review of 24 different appraisal tools used to evaluate CPG found AGREE had the most potential to serve as a critical appraisal tool for guidelines.[18] The use of the AGREE II instrument is

now widespread in published scientific research and it has also been adopted by major health bodies such as the World Health Organization (WHO) in their evaluation of CPG.[19]

This study aimed to evaluate the quality of cardiac CPG published over the last 10 years using the AGREE II instrument.

Methods

Review protocol

This study was performed in accordance with the guidelines from the preferred reporting items for systematic reviews and meta-analyses (PRISMA).

Information sources

This study was based on CPG published in seven journals. The authors of the study selected peer-reviewed journals in cardiology and cardiac surgery which have a history of publishing CPG for clinicians within the United States and Europe. These journals were *Annals of Thoracic Surgery (ATS)*, *Circulation*, *European Heart Journal (EHJ)*, *European Journal Of Cardiothoracic Surgery (EJCTS)*, *Heart*, *Journal Of The American College of Cardiology (JACC)* and *Journal of Thoracic and Cardiovascular Surgery (JTCVS)*. CPG published in these journals from January 2001 to May 2011 were identified using computerized searches on MEDLINE, SCOPUS and Google Scholar.

Eligibility criteria and search strategy

In order to avoid subjective interpretation for what constituted a guideline, the authors of the study agreed on an eligibility criteria for the purpose of the computerized search. Articles were considered guidelines if their title included the following terms: 'guidelines', 'recommendations', 'guidance', 'policy statement', 'consensus statement' or 'position paper'. These terms were paired with: 'cardiology', 'cardiovascular' and 'cardiac'. Guidelines were included only if they had a focus on preventative and/or therapeutic intervention. Guidelines with a focus on training, legal issues, epidemiology and research methods were excluded. Furthermore, editorial or correspondence articles that summarized organizational CPG were excluded.

Data extraction

The following information was extracted from each guideline: year of publication, journal(s) of publication, number of authors, organization(s) responsible for guideline, specialties involved in producing the guideline, region(s) of guideline development, funding for guideline development, guideline focus and whether it was an update of a previous CPG. Other information pertaining to guideline format included: the presence of a structured abstract, the description of a systematic literature review, the use of an evidence grade approach such as the American Heart Association Levels of Evidence and Grading of Recommendation,[20] an admission of a lack of evidence within the guideline, the declaration of conflicts of interests and the types of conflicts of interests declared.

Assessment of quality of guidelines

The AGREE II instrument is a tool used to assess the methodological quality of CPG.[16] The assessor must respond to 23 questions using a scale of 1 for 'strongly disagree' to 7 for 'strongly agree', based on examples and instructions described in the AGREE II manual.[17] AGREE II has six domains which are used to assess guideline quality; scope and purpose of the guideline, stakeholder involvement, rigour of development, clarity of presentation and editorial independence. Assessment of all 101 guidelines was performed by two of the authors (SS and AK). The scores of both authors were used to calculate an average for each domain and these scores were expressed as a percentage. A Spearman's correlation was used to assess for inter-rater reliability.

Data analysis

Data were analysed in SPSS 18.0 (SPSS Inc, Chicago, IL). The normality of the data was assessed. A comparison of guideline characteristics between 2001 to 2005 and 2006 to 2011 was performed using a chi-squared test (or a Fisher's exact test if data within one field were less than the value 5). These statistical tests were also performed to examine the effect of region, journal and specialty of authorship on guideline characteristics. AGREE II scores were compared taking into account study characteristics and average AGREE II scores were compared between two different time periods.

Results

Guideline characteristics

Two authors (SS and VP) performed the search based on the stated algorithm and identified 147 articles. Thirty guidelines were excluded because they were

Table 1. Chart demonstrating the number of guidelines identified in each journal over the study period.

Year of publication	Circulation	Journal of the American College of Cardiology	Heart	European Heart Journal	European Journal of Cardiothoracic Surgery	Annals of Thoracic Surgery	Journal of Thoracic and Cardiovascular Surgery
2001	1	1	1	2	0	0	0
2002	4	4	0	1	0	0	0
2003	4	3	0	3	0	0	0
2004	2	2	1	3	0	1	0
2005	1	1	2	5	1	3	1
2006	5	5	0	2	1	1	0
2007	7	5	2	5	0	2	0
2008	9	7	1	2	2	0	0
2009	8	6	0	2	1	1	0
2010	9	5	3	1	1	0	0
2011	4	3	0	0	0	0	0
Total	**54**	**42**	**10**	**26**	**6**	**8**	**1**

duplicates, five were excluded because they focused on research methods and training, and 11 were excluded because they were editorials on organizational guidelines. In total, 101 CPG met the inclusion criteria. Forty-five of the guidelines that were identified were published in more than one of seven journals. These guidelines were examined independently by the senior author (TA), and were all found to be eligible for inclusion in the study.

Circulation published the highest number of guidelines ($n = 54$) (Table 1). The *Journal of Thoracic and Cardiovascular Surgery* produced the fewest guidelines ($n = 1$).

The study period was split into two time periods; 'early' (2001–2005) and 'late' (2006–2011). In total, 35 guidelines were published in the early group and 66 guidelines were published in the late group. Table 2 illustrates the frequencies of guideline characteristics over the two time periods.

Reporting conflicts of interests

In the 'early' study period, none of the guidelines published their conflict of interests, while in the 'late' period 69.7% of the guidelines published them ($p < 0.001$). Guidelines originating from the United States were more likely to report on conflicts of interests ($p < 0.001$). Guidelines developed in Europe ($n = 34$) were less likely to report on conflicts of interests ($p < 0.001$).

Evaluation of guideline quality using AGREE II assessment tool

A comparison of overall AGREE II scores between the two raters using Spearman's correlation was 0.965, indicative of strong inter-rater reliability. Assessment of individual domains is presented in Table 3. 'Scope and purpose' (median score of 86.1%) and 'clarity of development' (median score of 83.3%) were the two domains within AGREE II that received the highest scores, 'applicability' (median score of 20.8%) and 'editorial independence' (median score of 33.3%), scored the lowest of all the AGREE II domains.

Comparison of individual AGREE II domain scores using an independent *t*-test found no significant change in scores from the first half of the study period compared to the second half. An overall AGREE II score for each guideline was not found to have a significant relationship with the geographic area of development, the association responsible for guideline development or the journal of publication. Guidelines that had a focus on a preventative intervention ($p = 0.003$) and that admitted to a lack of evidence within the body of the text ($p = 0.002$) had

Table 2. Table showing differences in number of guidelines and guideline characteristics over the study period.

	2001–2005 $n = 35$	2006–2011 $n = 66$	Total	p value (Chi Squared/ Fisher's Exact Test*)
Journal				
Circulation	p12	42	54	**0.005**
Journal of the American College of Cardiology	11	31	42	0.132
Annals of Thoracic Surgery	4	4	8	0.443*
European Heart Journal	14	12	26	**0.017**
European Journal of Cardiothoracic Surgery	1	5	6	0.662 *
Heart	4	6	10	0.735*
Organization				
American Heart Association	13	41	54	0.017
American College of Cardiology	11	32	43	0.099
European Society of Cardiology	14	14	27	**0.028**
European Society of Cardiothoracic Surgeons	1	5	6	0.662*
Society of Thoracic Surgeons	5	11	16	0.755
National Institute for Health and Clinical Excellence	0	5	5	0.161*
Region of development				
USA	18	47	65	**0.048**
Europe	17	17	34	**0.021**
UK	3	6	9	1*
Other	1	1	2	1*
Funding				
Government	1	7	8	0.256*
Medical Association	34	61	95	0.662*
Private/Industry	1	0	1	0.347*
Focus				
Prevention	2	20	22	**0.004**
Therapy intervention	13	17	30	0.206
Both	20	29	49	0.233
Other guideline characteristics				
Structured abstract	8	20	28	0.426
Based on a previous guideline	11	24	35	0.62

(continued)

Table 2. Continued.

	2001–2005 $n = 35$	2006–2011 $n = 66$	Total	p value (Chi Squared/ Fisher's Exact Test*)
Systematic review performed	1	19	20	**0.002**
Use of an evidence grade approach	24	48	72	0.66
Use of AHA approach	22	45	67	0.59
Lack of evidence admitted	27	50	77	0.876
Appropriateness criteria guideline	0	5	5	0.161*
Conflict of interests declared	0	46	46	**0**

Note: *Fischer's exact test was used to assess significance instead of a chi-squared test.

Table 3. Descriptive statistics of performance of guidelines across all six AGREE II domains.

AGREE II domain	Mean	Median	Standard deviation	Variance	Range	First quartile	Second quartile	Third quartile
Scope and purpose (%)	85.05	86.1	8.36	69.93	61.10–100	80.6	86.1	91.7
Stakeholder involvement (%)	58.5	58.3	6.67	44.52	39.90–80.60	55.6	58.3	63.9
Rigour of development (%)	45.99	42.3	15.7	246.49	16.70–83.30	35.4	42.7	58.3
Clarity of development (%)	81.77	83.3	10.17	103.52	27.80–100	77.8	83.3	88.9
Applicability (%)	22.35	20.8	11.09	122.99	4.20–54.20	14.6	20.8	30.22
Editorial independence (%)	28.84	33.3	24.47	598.83	0–62.50	0	33.3	54.2

higher overall AGREE II scores. When individual AGREE II domains were examined in relation to these three guideline demographics, guidelines published in *Heart* were found to have a higher score for 'Scope and purpose' (median score of 88.90, $p = 0.04$) compared to guidelines published in other journals. 'Applicability' scores were also higher in guidelines developed by the American Heart Association (AHA) (median score of 24.50, $p = 0.03$). The use of an evidence grade approach, description of a systematic review and declaring conflicts of interests did not significantly improve individual AGREE II scores. Further comparison of other guideline characteristics that were extracted from each guideline any significant relationships when compared to individual AGREE II domains.

Discussion

This study has found that over the last 10 years there has been marked variability in the quality of guidelines published in cardiac clinical practice. Furthermore, there was no evidence of improvement in overall quality or improvement within the individual AGREE II domains. Our findings are concordant with the assessments of the guidelines across multiple spheres of clinical medicine:[13] guidelines do not adhere to established methodological standards and there was a great need for further improvement in guideline development. The failure to demonstrate an improvement in quality over the course of the study period despite recommendations raises significant concerns and warrants repeated future assessment to ensure that guideline developers take heed

of the identified problems. Guideline readers and practising clinicians must take steps to assess guideline quality and this can be easily performed using the AGREE II tool.

Geographical variations in guideline production

A significant increase in the number of guidelines produced within the United States and by the AHA was seen in the 'late' study period. There was a reciprocal fall in the number of guidelines produced in Europe and by the European Society of Cardiology. However, changes in the quantity of guidelines per geographical area are not reflected in the change in quality: there was no association between the overall AGREE II scores or individual AGREE II domains, and the region of development. In view of this, regional differences in the number of guidelines produced warrants further examination. While some European countries need assistance to improve their productivity in biomedical research,[21] the actual gap between Europe and the USA in production of scientific research is very small.[22] Therefore, if both regions appear to be equal in the production of scientific research, the variation in guideline production in clinical medicine is of interest. The larger volume of guidelines produced in the United States may reflect how the regional medicolegal pressures affect the volume of guidance produced by their healthcare bodies. Medical organizations within the United States may have a higher commitment to updating older guidelines, and this contributes significantly to their increased guideline production. Conversely, publicly funded bodies in Europe may be unable to financially support regular guideline publication.

Changes in clinical evidence and healthcare resources are key reasons to regularly evaluate and update CPG.[23] It is therefore important that European cardiovascular groups recognize the need to keep up with their American colleagues within this field so that healthcare professions within their region can rely on their guidance, with the knowledge that it is current and up to date.

Systematic reviews improve the quality of guidelines

The increase in the latter half of the study of the number of guidelines that described a systematic literature review was an important finding. The inclusion of a systematic review was not found to improve the individual AGREE II domains. Although 'Rigour of development' is a domain which includes the use of a systematic review within its assessment of a guideline, there are a number of other factors which are included as well, and therefore the authors of this study were not surprised that a statistically significant relationship was not found between the two variables. Detailed literature is available to help guideline writers to perform a thorough systematic review,[24,25] and their inclusion within a guideline adds to the methodological quality and ensures scientific validity.[13] Furthermore, the Institute of Medicine's (IOM)[1] recommendations on guidelines state that in order for a guideline to be considered trustworthy it should be based on a systematic review. It is anticipated that systematic reviews become a gold standard for guidelines and there use is included in AGREE II as a marker of quality in rigour of development. The authors of this study would support recommendations of the IOS and AGREE II that and their use should be encouraged in guideline development.

Conflicts of interest: room for improvement

Although reporting of conflicts of interests also improved in the second half of the study period, European guidelines were less likely to report on conflicts of interests. The failure of published guidelines to disclose conflicts of interests has previously been described.[14] Despite tighter regulation, the pharmaceutical and medical-device industries have a strong financial need to influence physicians.[26] It is commonplace for many authors to have some interaction with industry, with many serving on the advisory boards of industry or have related patents.[27] Such interactions by authors may introduce bias into guidelines, whether conscious or not. Guideline readers must be made aware of this potential for bias for each guideline. It is not enough to make conflicts of interest available in a separate document, as this provides an additional step that hinders awareness of the influence of industry.

Variability in guidelines: identifying areas that require improvement

Examination of the individual AGREE II domains found that 'Scope and purpose' and 'Clarity of presentation' were the two areas which had the highest scores. Although these results were encouraging, 'Developmental rigour' which is the largest domain in AGREE II did not perform as well. One, the key components within this domain are the way in which recommendations are formulated and presented within a guideline. The AHA and American College of Cardiology (ACC) classification system[20] was the most commonly adopted approach and was used in 66% of guidelines. None of the guidelines used the Grading of Recommendations, Assessment,

Development and Evaluation tool (GRADE).[28] There is emerging consensus that GRADE should be adopted for guidelines in all areas of clinical medicine and its use seems to becoming increasingly common.[29–31] The benefit of a single classification system for all specialties in medicine is that interpretation of guideline recommendations may become more straightforward.

'Applicability' within AGREE II assesses whether a guideline provides advice for implementing its recommendations and addresses the costs and resource required. Overall, 'applicability' was found to have the lowest of all six domains suggesting guidelines fail to consider and advise on how to apply key recommendations within clinical practice. Further development of guidelines should consider this factor and address its importance so that healthcare professionals may adopt recommended practice more easily.

The value of AGREE II for guideline developers and readers

Use of AGREE II by guideline developers could help them identify areas within the guideline requiring improvement before publication. Furthermore, guideline readers may benefit from the publication of AGREE II scores alongside the guideline to objectively assess its quality and thus the strength of its recommendations. Indeed, AGREE II scores may be scientifically more beneficial than editorials accompanying guidelines.

Readers may attribute particular importance to guidelines based upon the journal of its publication or the organization responsible for its development; however, this study demonstrates these factors have no effect on the citation of randomized evidence or the overall quality of the guideline. Thus, caution is required for healthcare professionals relying solely upon 'brand identity' when considering recommendations of a guideline.

Study limitations

There were several limitations to this study. Firstly, the inclusion of only seven journals in cardiac medicine and surgery meant that the findings in this study are restricted to the included journals. However, those journals were chosen because they were viewed to be prestigious with a history of publication of clinical guidelines. The authors of this study believe that problems that were identified within this study have a particular impact on healthcare practice within the USA, Europe and the UK because these regions are likely to rely on the guidance published in these journals.

A second limitation was the reliance on computerized searches to identify guidelines and references. Although a robust set of searching criteria was formulated and tested prior to guideline identification, there is a possibility that some guidelines were missed. A paper search may have resulted in more human error. Third, the assessment of reporting on conflicts of interest did not distinguish between guidelines that obtained disclosures but did not publish them and guidelines that failed to obtain disclosures. Although the guideline development bodies or journals may have internally scrutinized the conflicts of interest, a failure to publish these implies that guideline reader cannot reflect on the impact that industry-related or individual conflicts of interest may have affected guideline development. Furthermore, the authors adopted this methodology because recommendations on guideline development require the publication of conflicts of interests within the body of the guideline.[17] Finally, the AGREE II tool has a broad range for the assessment of an individual guideline component. This makes the decision-making process for the rater more subjective in nature. Despite this, the correlation between the raters was strong suggesting AGREE II overcomes this potential bias. This is consistent with existing research that indicates supports the validity of AGREE II as an assessment tool for specialists across all medical specialties and at different levels of seniority.[32]

Conclusions

There is considerable variability in the quality of cardiac CPG. Guideline development should incorporate the use of validated guideline assessment tools such as AGREE II which may improve the evidence base and methodology of guidelines. Clinicians can use such assessment tools to appraise guidelines, which may help them decide which recommendations they should adopt in their clinical practice.

Declarations

Competing interests: None declared

Funding: None declared

Ethical approval: Not applicable

Guarantor: TA

Contributorship: SS, VP and TA were responsible for study design, data collection and analysis. Furthermore, they were responsible for drafting and revising the original manuscript. AK and SN were responsible for data collection, analysis and revision of the paper. AD, JK, IM and JC were responsible for study conception and design of the study, as well as revision of the paper. All authors gave final approval of the version to be published.

Provenance: Submitted; peer reviewed by Ami Banerjee and Petros Skapinakis.

References

1. *Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines: Clinical Practice Guidelines We Can Trust*. Washington, DC: Institute of Medicine, 2011.
2. Grimshaw JM and Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; 342: 1317–22.
3. Jackson R and Feder G. Guidelines for clinical guidelines. *BMJ* 1998; 317: 427–8.
4. Cook D and Giacomini M. The trials and tribulations of clinical practice guidelines. *JAMA* 1999; 281: 1950–1.
5. Heffner JE. Does evidence-based medicine help the development of clinical practice guidelines? *Chest* 1998; 113: 172–8.
6. Tricoci P, Allen JM, Kramer JM, et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009; 301: 831–41.
7. Hadorn DC, Baker D, Hodges JS, et al. Rating the quality of evidence for clinical practice guidelines. *J Clin Epidemiol* 1996; 49: 749–54.
8. Giannakakis IA, Haidich AB, Contopoulos-Ioannidis DG, et al. Citation of randomized evidence in support of guidelines of therapeutic and preventive interventions. *J Clin Epidemiol* 2002; 55: 545–55.
9. Detsky AS. Sources of bias for authors of clinical practice guidelines. *CMAJ* 2006; 175: 1033–5.
10. Thomson R, McElroy H and Sudlow M. Guidelines on anticoagulant treatment in atrial fibrillation in Great Britain: variation in content and implications for treatment. *BMJ* 1998; 316: 509–13.
11. American Medical Association. *Office of Quality Assurance. Attributes to Guide the Development and Evaluation of Practice Parameters*. Chicago, IL: American Medical Association, 1990.
12. Eddy D. *A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach*. Philadelphia, PA: American College of Physicians, 1992.
13. Shaneyfelt TM, Mayo-Smith MF and Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; 281: 1900–5.
14. Papanikolaou GN, Baltogianni MS, Contopoulos-Ioannidis DG, et al. Reporting of conflicts of interest in guidelines of preventive and therapeutic interventions. *BMC Med Res Methodol* 2001; 1: 3.
15. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010; 182: 839–42.
16. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *J Clin Epidemiol* 2010; 63: 1308–11.
17. *AGREE Manual*. See http://www.agreetrust.org (last checked 20 March 2013).
18. Vlayen J, Aertgeerts B, Hannes K, et al. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care* 2005; 17: 235–42.
19. World Health Organization. *The Reproductive Health Library (2011)*. Geneva: WHO, 2011. See http://apps.who.int/rhl/guidelines/en/index.html (last checked 20 March 2013).
20. *American Heart Association Levels of Evidence and Grading of Recommendation*. See http://www.aha.org (last checked 20 March 2013).
21. Soteriades ES and Falagas ME. Comparison of amount of biomedical research originating from the European Union and the United States. *BMJ* 2005; 331: 192–4.
22. EU eliminates citation gap with America. *Nature* 1997;**387**:537.
23. Shekelle P, Eccles MP, Grimshaw JM, et al. When should clinical guidelines be updated? *BMJ* 2001; 323: 155–7.
24. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997; 127: 380–7.
25. Meade MO and Richardson WS. Selecting and appraising studies for a systematic review. *Ann Intern Med* 1997; 127: 531–7.
26. Wazana A. Physicians and the pharmaceutical industry: is a gift ever just a gift? *JAMA* 2000; 283: 373–80.
27. Krimsky S, Rothenberg LS, Stott P, et al. Scientific journals and their authors' financial interests: a pilot study. *Psychother Psychosom* 1998; 67: 194–201.
28. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924–6.
29. Canfield SE and Dahm P. Rating the quality of evidence and the strength of recommendations using GRADE. *World J Urol* 2011; 29: 311–7..
30. Onate-Ocana LF and Ochoa-Carrillo FJ. The GRADE system for classification of the level of evidence and grade of recommendations in clinical guideline reports. *Cir Cir* 2009; 77: 417–9.
31. Schunemann HJ, Oxman AD, Brozek J, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evid Based Med* 2008; 13: 162–3.
32. Brouwers MC, Kho ME, Browman GP, et al. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ* 2010; 182: 472–8.