International Journal of Technology Assessment in Health Care, **16:4** (2000), 1024–1038.

A COMPARISON OF CLINICAL PRACTICE GUIDELINE APPRAISAL INSTRUMENTS

Ian D. Graham
Ottawa Hospital – Civic Campus

Lisa A. Calder
University of Western Ontario

Paul C. Hébert
University of Ottawa

Anne O. Carter
United Arab Emirates University

Jacqueline M. Tetroe
Ottawa Hospital – Civic Campus

Abstract

Objective: To identify and compare clinical practice guideline appraisal instruments.
Methods: Appraisal instruments, defined as instruments intended to be used for guideline evaluation, were identified by searching MEDLINE (1966–99) using the Medical Subject Heading (MeSH) practice guidelines, reviewing bibliographies of the retrieved articles, and contacting authors of guideline appraisal instruments. Two reviewers independently examined the questions/statements from all the instruments and thematically grouped them. The 44 groupings were collapsed into 10 guideline attributes. Using the items, two reviewers independently undertook a content analysis of the instruments.
Results: Fifteen instruments were identified, and two were excluded because they were not focused on evaluation. All instruments were developed after 1992 and contained 8 to 142 questions/statements. Of the 44 items used for the content analysis, the number of items covered by each instrument ranged from 6 to 34. Only the instrument by Cluzeau and colleagues included at least one item for each of the 10 attributes, and it addressed 28 of the 44 items. This instrument and that of Shaneyfelt et al. are the only instruments that have so far been validated.
Conclusions: A comprehensive, concise, and valid instrument could help users systematically judge the quality and utility of clinical practice guidelines. The current instruments vary widely in length and comprehensiveness. There is insufficient evidence to support the exclusive use of any one instrument, although the Cluzeau instrument has received the greatest evaluation. More research is required on the reliability and validity of existing guideline appraisal instruments before any one instrument can become widely adopted.

Keywords: Practice guidelines, Knowledge, Attitudes, Practice, Quality of health care

Ian Graham and Paul Hébert are Career Scientists with the Ontario Ministry of Health. IG was a Medical Research Council of Canada Postdoctoral Health Research Fellow when the study was initiated. We would like to thank Laura McAuley and Christine Niles for their assistance in preparing this manuscript.

Clinical practice guidelines are systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances (15). When acted upon, they have been shown to have the potential to improve both the process of care and patient health outcomes (9;10;27;36). The beneficial effects of guidelines will not result, however, unless well-developed and valid guidelines are implemented by clinicians and/or policy makers.

It is estimated that some 2,500 guidelines are already in existence (36). With the exponential growth in guideline development, clinicians are increasingly being confronted with differing and sometimes contradictory disease-specific guidelines (7;29;31). In one study from Britain (31), the recommendations from 20 practice guidelines on anticoagulation treatment in atrial fibrillation were applied to 100 consecutive patients. Depending on the guideline, anticoagulant treatment would have been recommended for 13% to 100% of the patients. This case study demonstrated how the application of differing practice guidelines for the same condition had major implications for clinical decision making as well as the quality of care patients received. The authors of the study attributed the variation found in the guidelines to their nonsystematic development. Others have also raised concerns about the quality of guidelines that are being developed (4;8;28;33;34).

The rapid rate of development of practice guidelines would appear to be a barrier to their implementation. Determining which guidelines are quality products worthy of use can be quite a daunting task for busy clinicians and policy makers. Some authors (14) have suggested that inability to critically appraise the quality of clinical practice guidelines has been a barrier to their use. Others (9;12;16) have proposed that if physicians are provided with instruments to systematically appraise guidelines, the adoption of high-quality and useful guidelines may be increased. At the same time, this approach might also be a prudent strategy to help overcome another barrier to guideline use—the perceived threat to clinicians' autonomy that guidelines pose for some physicians. Indeed, when clinicians uncritically accept and apply clinical practice guidelines, their decision-making autonomy is undermined, because guidelines are intended only to guide practice and still must be applied judiciously. However, if clinicians have the tools to critically appraise clinical practice guidelines and can assess their quality and utility for themselves, they retain independence over how best to treat individual patients. Other users of practice guideline appraisal instruments are healthcare administrators of paying agencies, either in government (in public healthcare systems) or in healthcare enterprises such as health maintenance organizations (HMOs) and insurance companies. If valid and reliable instruments can be agreed to by all parties (payers, providers, and patients), then the quality and feasibility of guidelines endorsed by policy makers can be improved. Approving the method by which the quality of guidelines is to be assessed should enhance the acceptance of endorsed guidelines by healthcare providers.

The Institute of Medicine (IOM) of the National Academy of Sciences in the United States has been interested in the quality of practice guideline since the early 1990s, when it delineated eight desirable attributes of guidelines. The attributes were intended to help guideline "users understand the elements of a sound guideline and to recognize good (or not-so-good) guidelines" (16). The IOM argued that each attribute affects the likelihood that guidelines will be perceived as trustworthy and usable or the probability that they will, if used, help achieve the desired health outcomes. As described by the IOM, the eight desirable attributes fell into two categories: four attributes related to the substantive content of the guideline (validity, reliability, clinical applicability, clinical flexibility), and four related to the process of guideline development or guideline presentation (clarity, multidisciplinary process, scheduled review, documentation). The desirable attributes for guidelines put forth by the IOM were later adopted by the Agency for Health Care Policy and Research (AHCPR), an agency legislated by the U.S. Congress to ". . . improve the quality,

appropriateness, and effectiveness of health services . . .” (22). Attempting to operationalize the desirable attributes of guidelines, the IOM put forth a provisional instrument for assessing practice guidelines (20). The IOM has never finalized this instrument, which is long (142 items) and has been found to be cumbersome and impractical to apply (5).

Interested in evaluating the state of the art and science of guideline appraisal, we identified and compared existing guideline appraisal instruments.

METHODS

Literature Search

To identify existing practice guideline appraisal instruments, we conducted a search of the English language literature indexed in the MEDLINE database (1966–99), using the Medical Subject Heading (MeSH) practice guidelines. A second search of the same database was conducted using the MeSH practice guidelines.pt (pt = publication type) and practice guidelines.tw (tw = text word). Manual searches of the retrieved articles' bibliographies were conducted, and articles from personal collections (PCH, AOC) were included. These documents were retrieved and reviewed. We then selected appraisal instruments for comparison if the authors stated or implied that their instrument could be used to help readers obtain information needed to assess the quality of guidelines. We also attempted to contact the developers of the identified appraisal instruments to determine whether they were aware of any other instruments that might have been missed.

Content Analysis

The content analysis was a two-stage process. The first stage involved generating the items to be used in comparing each instrument by compiling a list of all questions/statements from each of the instruments. Two reviewers (LAC and IDG) independently examined the list and grouped common questions/statements. These groupings became the “items” for the content analysis. For example, all questions/statements related to whether the purpose of the guideline was stated were grouped together and given an item label; in this case the label was “purpose.” The reviewers then met and compared their two lists of items. At this point, consensus was reached on the items and the labels to be used to describe each.

In total, 44 items were generated from the list of 394 questions/statements compiled from 12 instruments. Only two questions/statements were assigned to an “other” category. Of the 44 items, 39 were more “objective” in nature and involved the guideline appraiser's examining the text of the guideline or the accompanying documentation for a statement or discussion of specific issues (for example, whether there was a description of how consensus about the recommendation was reached). The remaining five items required the appraiser to make a “subjective” assessment of the guideline with respect to the issues of its clinical flexibility, its ambiguity of wording, its presentation, its ease of use, and the existence of any conflicts of interest among the guideline developers.

Next, two reviewers (IDG, JMT) jointly collapsed these items into broad common categories. These categories were largely derived from the IOM's attributes of guidelines and from the work of Cluzeau et al. We labeled these groupings of common items “guideline attributes” (e.g., validity, clinical applicability). We grouped the 44 individual items into 11 major guideline attributes. The attributes reflected issues related to the methodologic rigor with which the guideline was developed (*validity, reliability/reproducibility*), the clinical content of the guideline (*clinical applicability, clinical flexibility*), the process of guideline development (*multidisciplinary process*), the presentation of the guideline (*clarity*), the currency and updating of the guideline (*scheduled review*), guideline dissemination (*dissemination*), the feasibility and implications of implementing the guideline (*implementation*), evaluation of the guideline's impact (*evaluation*), and a residual category that was used for

items that did not fall into any of the preceding guideline dimensions (*other*). The appraisal items comprising each guideline attribute used for the content analysis, along with their definitions, are presented in Table 1.

The second stage of the analysis involved two reviewers (LAC and IDG) independently examining each instrument for the presence of the items identified during the first stage. The reviewers then met and compared the results of their content analysis of each instrument. Disagreement was resolved through consensus. A third reviewer (JMT) confirmed the content analysis by once again examining the questions/statements of each instrument and the items to which they had been assigned. Questionable assignment of items was resolved by consensus. A list of the actual instrument questions/statements assigned to each item is available from the authors upon request.

RESULTS

General Description of Instruments

A total of 15 possible practice guideline appraisal instruments were identified by the search process. Two were identified by one of the instrument developers we contacted (23;24). Two (2;30) were eventually excluded because they were designed as guiding principles for guideline development and implementation rather than instruments for evaluating guidelines. A description of the 13 instruments is presented in Table 2. All were published after 1992. The instruments originated from five different countries: Australia (17;34), England (3), Scotland (24), the United States (13;20;21;25;26;35), and Canada (1;12;13;23). The number of questions/statements in the appraisal instruments range from as few as 8 to as many as 142. The target users included guideline developers (3;5;12;17;20), clinicians (generalists and specialists) (12;13;21;25;26;35), public health organizations (24;34), governmental agencies (34), and anyone interested in evaluating guidelines (1;12;17;23;26). In terms of the stated purpose of the instrument, three were intended to help readers decide whether they should use a guideline (13;25;35), two were designed to promote systematic guideline development (3;20), two were intended to evaluate guideline validity (3;17), five were for quality appraisal (1;13;24;26;34), and one was designed to evaluate the clinical practice guideline development process (23). Eight of the instruments were stated to be based on guideline attributes put forth by the IOM (3;13;20;21;23;24;26;34). Only two (1;17) referenced instruments other than the one produced by the IOM. The number of references to the literature cited in the documentation accompanying the instruments ranged from none to 91. All but three instruments were published in the peer-reviewed literature (17;23;24). Only two instruments permitted any scaling of the items (3;26); these same instruments were the only two that have been subjected to any sort of validation studies (4;18;26).

Content Analysis

Table 3 presents the comparison of each instrument against the items generated during the first stage of the content analysis. The instruments varied considerably in terms of the number of items each addressed. Of a total of 44 possible items, the number covered by each instrument ranged from 6 (21) to 34 (20). The items common to the most instruments were: a statement of the patient population (11 instruments); a statement of the guideline topic, independent review of the guideline (10 instruments each); and a statement about the recommendation and the evidence for it, a statement of the outcomes, and composition of the guideline development group (nine instruments each).

The guideline attributes covered by the most instruments were validity, reliability/reproducibility, and clinical applicability of guidelines (12 instruments each). The next most common attribute was scheduled review (11 instruments), followed by clinical flexibility and multidisciplinary process of guidelines (10 instruments each). The attribute assessed

Table 1. Definitions of Items Used in the Content Analysis of Clinical Practice Guideline (CPG) Appraisal Instruments

Dimension	Item label	Definition
Validity	Decision making: How consensus was reached	Method(s) used to reach consensus about CPG recommendation (e.g., open discussion, Rand/ Delphi technique); role of values
	Decision making: How recommendations were made	Method(s) used in formulating recommendations
	Evidence collection	How the evidence was obtained
	Literature search	How the literature was searched, including search strategy
	Sources of evidence	Sources of evidence, such as text books, periodical literature
	References cited	References for the evidence upon which the CPG was based
	Literature selection	The criteria used to include/exclude literature from the data synthesis
	Evaluation of evidence	How the evidence was graded, which may or may not include a statement about the strength of evidence
	Data synthesis	Method(s) by which the evidence was synthesized (e.g., meta-analysis, systematic review)
	Recommendations and evidence for them	Recommendations consistent with each other and the evidence used to support them
	Links strength of evidence to recommendation	Links strengths of evidence to recommendation
	Health benefits	Expected health benefits of CPG
	Harms, risks	Potential harms or risks of CPG
	Costs	Economic and other cost outcomes of CPG
	Outcomes stated	Outcomes expected to result from CPG
	Other CPGs	The existence of other CPGs relevant to CPG topic
	Alternatives	Alternative interventions to those recommended or dealt with by the CPG to deal with topic
Reliability/ reproducibility	Independent review	CPG sent to experts not involved in its development for review
	Pilot/pretesting	CPG pilot or pretested in clinical setting prior to dissemination
	Documentation	Process of guideline development documented
Clinical applicability	Purpose	The goal or objective of the CPG
	Rationale	The rationale of or reason for the CPG
	Guideline topic	Guideline topic
	Patient population	The patient population(s) for whom the CPG is intended
	Provider population	The group(s) of healthcare providers to whom the guideline is directed or who should use the CPG
Clinical flexibility	Exceptions/flexibility[a]	Flexibility in the application of the CPG, or situations in which CPGs may not apply
	Patient preferences considered	Whether patient choices and/or views considered
Clarity	Unambiguous[a]	CPG is clearly worded
	Presentation[a]	CPG presentation is user friendly
	Ease of use[a]	CPG can be used in a straightforward manner
	Structured abstract	Structured abstract provided

(*Continued*)

Table 1. (*Continued*)

Dimension	Item label	Definition
Scheduled review	Date of issue of CPG	Date of issue of CPG
	Expiry date/ scheduled review	Date CPG no longer valid or is scheduled for review
Multidisciplinary Process	Conflict of interest[a]	Consideration of any bias, conflicts of interest, potential bias, or potential conflicts of interest related to the individuals developing the CPG; or offers information that the CPG appraiser can use to infer actual or potential conflict of interest
	Funding (and related bias)	Sources of funding
	Composition of guideline development group	The individuals and/or disciplines, occupations, or organizations they represented in the group who developed the CPG (e.g., surgeons, nurses, patient representatives)
	Feedback	Developers obtained response to CPG from potential users
	Guideline development organization	The organization or group who developed the CPG
	Endorsers	Endorsement of CPG by official bodies
Dissemination	Dissemination	How the CPG is to be distributed to intended users
Implementation	Implementation	Strategies (in addition to dissemination to promote use of CPG)
	Policy and administrative implications (feasibility)	Policy and administrative implications of using CPG
Evaluation	Evaluation	How the CPG is to be evaluated once it has been implemented (e.g., health, economic, patient satisfaction, provider satisfaction, outcomes)
Other	Other	Unique codings

All the items with the exception of those indicated by footnote *a* required that the items be stated in the text of the CPG.

[a] Denotes subjective items that required the individual appraising the CPG to make a personal decision.

by the fewest instruments was guideline dissemination. Only the Cluzeau instrument (3) had an item related to this.

The extent to which the instruments had items that fell within each of the 10 guideline attributes (excluding the "other" category) also differed. Only the Cluzeau instrument included at least one item that related to each of the guideline attributes. Two instruments (23;24) had items for nine guideline attributes, and three had items for seven of the attributes (20;25;34). The fewest number of attributes covered by an instrument was five (13;17;35).

The instruments also differed in the degree to which each guideline attribute was assessed (Table 4). Only one instrument included every item for the attribute of clinical validity (26), and no instrument mentioned every item included for the attribute of validity, reliability/reproducibility, or multidisciplinary process. Furthermore, the total number of items addressed within each attribute differed by instrument. For example, of the 17 possible items comprising the attribute of validity, the number covered by each instrument ranged from zero (21) to 15 (20).

DISCUSSION

In this study we were able to locate 13 instruments for evaluating clinical practice guidelines. A review of the background of these instruments and a content analysis of each revealed that

Table 2. Characteristics of Guideline Appraisal Instruments

Author	Title	Date	Target user	Purpose	Basis of development/ testing	No. of refs.	Published in peer-reviewed literature	Validation	Scaling of questions	No. of questions
IOM (20)	A provisional tool for assessing clinical practice guidelines	1992	Clinicians, guideline developers, researchers	To provide an explicit method for examining the soundness of guidelines and encouraging their systematic development	IOM committee and outside consultant operationalized 8 conceptual attributes of guidelines developed by IOM committee Internal IOM review process Piloted on guidelines with 59 clinicians Instrument reviewed by 7 external anonymous experts Authors recommend pretesting by other groups	2	Yes	Not stated	No	142
Hayward et al. (12)	More informative abstracts of articles describing clinical practice guidelines	1993	Physicians, guideline developers and implementers	To propose a structured abstract that could enhance readers' ability to appraise guidelines	Emerging principles of guideline development and evaluation Reviewed and pilot tested by guideline developers and evaluators, implementers, and practicing clinicians	35	Yes	Not stated	No	10
Selker (25)	Criteria for adoption in practice of medical practice guidelines	1993	Practitioners	To develop useful criteria for adoption of guidelines for clinical practice	Not stated	20	Yes	Not stated	No	10

Hayward et al. (13)	Users guides to the medical literature, VIII: How to use clinical practice guidelines, A: Are the recommendations valid?	1995	Clinicians	To help clinicians decide whether to use a guideline in formulating their own clinical policy	IOM work, unpublished work of S.H. Woolf	30	Yes	Not stated	No	13
Mendelson (21)	The development and meaning of appropriateness guidelines	1995	Radiologists	To develop appropriateness criteria for guidelines	AHCPR attributes (=IOM attributes)	13	Yes	Not stated	No	8
Woolf (35)	Practice guidelines: What the family physician should know	1995	Family physicians	To review key information family physicians should have in order to evaluate and use guidelines effectively	Not stated	50	Yes	Not stated	No	10
SIGN (24)	Clinical guidelines: Criteria for appraisal for national use	1995	Scottish Intercollegiate Guidelines Network (SIGN)	To develop criteria by which SIGN will appraise guidelines recommended for national use in Scotland	Clinical Resource and Audit Group (CRAG), AHCPR, IOM	6	No	Not stated	No	50
Cluzeau et al. (3)	Appraisal instrument for clinical guidelines	1996	Guideline developers	To appraise validity and content of existing guidelines and to help establish a framework for systematic development of new guidelines	IOM appraisal instrument; pilot-tested on 31 guidelines and revised	0	Yes	Yes, reliability and validity assessed by having 120 reviewers use the instrument to evaluate 60 guidelines (see ref. 28)	Yes, partially by section only	37

(*Continued*)

Table 2. (*Continued*)

Author	Title	Date	Target user	Purpose	Basis of development/ testing	No. of refs.	Published in peer-reviewed literature	Validation	Scaling of questions	No. of questions
Liddle et al. (17)	Method for evaluating research guideline evidence	1996	Guideline reviewers and developers	To allow reviewers to assess whether guidelines or recommendations are valid and likely to benefit a population	IOM, consultation with epidemiologists, Cochrane Collaboration and clinicians on expert panel (details not specified), pilot-tested	44	No	Not stated	No	14
Ward and Grieco (34)	Why we need guidelines for guidelines	1996	Clinical colleges, faculties, federal & state health departments, national & state NGOs	To appraise quality of guidelines produced in Australia	IOM appraisal instrument; instrument used to evaluate 34 guidelines	10	Yes	No	No	18
Savoie et al. (23)	Critical appraisal criteria for clinical practice guidelines	1996	Researchers (experts in critical appraisal of various disciplines) and guideline developers	To review and evaluate the processes used in developing guidelines and to determine the extent to which guidelines are evidence-based	IOM and AHCPR work; instrument used on some guidelines	13	No	Not stated	No	33
Calder et al. (1)	Review of published recommendations for transfusion of allogeneic red cells blood and plasma	1997	Guideline evaluators	To evaluate existing guidelines for transfusion of allogeneic red blood cells and plasma	See key articles by Woolf, references 21 and 24; used to evaluate 17 guidelines	91	Yes	No	No	24
Shaneyfelt et al. (26)	Are guidelines following guidelines? The methodologic quality of clinical practice guidelines in the peer-reviewed medical literature	1998	Clinicians, researchers	To assess the adherence of published CPGs to established methodologic standards for CPGs	Literature review; dialogue with representatives from organizations that develop, implement, evaluate, or use CPGs; items pilot-tested at 3 national workshops; pretested by authors	31	Yes	Endorsed as comprehensive and valued by CPG experts; content validity procedure decribed	Yes, by section and total	25

Table 3. Comparison of Instruments by Guideline Attribute

Attribute	Total possible items	IOM (20)	Hayward et al. (12)	Selker (25)	Hayward et al. (13)	Mendelson (21)	Woolf (35)	SIGN (24)	Cluzeau et al. (3)	Liddle et al. (17)	Ward and Grieco (34)	Savoie et al. (23)	Calder et al. (1)	Shaneyfelt et al. (26)
Clinical applicability	5	4	4	0	3	1	0	3	4	3	2	2	3	5
Clinical flexibility	2	2	0	1	0	1	0	1	2	1	1	1	1	2
Reliability/ reproducibility	3	2	2	1	2	1	2	2	2	0	1	2	2	1
Validity	17	15	12	1	7	0	4	10	9	8	7	9	8	14
Clarity	4	4	1	1	0	1	0	4	1	0	3	2	0	0
Scheduled review	2	2	0	1	1	0	1	1	1	1	2	1	1	1
Multidisciplinary process	6	4	4	1	0	1	1	5	5	0	0	3	1	1
Dissemination	1	0	0	0	0	0	0	0	1	0	0	0	0	0
Implementation	2	0	0	0	1	0	1	2	1	1	1	1	0	0
Evaluation	1	0	0	1	0	1	0	1	1	0	0	1	1	0
Other	1	1	0	0	0	0	0	0	0	0	0	1	0	0
Total no. of items	44	34	23	7	14	6	9	30	27	14	17	23	17	23

Table 4. Comparison of Instruments by Guideline Attribute

Attribute	Total possible items	IOM (20)	Hayward (13)	Selker (25)	Hayward (11)	Mendelson (21)	Woolf (35)	SIGN (24)	Cluzeau (3)	MERGE (17)	Ward (34)	BCOHTA (23)	Calder (1)	Shaneyfelt (6)
Clinical applicability	5	4	4	0	3	1	0	3	4	3	2	2	3	5
Clinical flexibility	2	2	0	1	0	1	0	1	2	1	1	1	1	2
Reliability/ reproducibility	3	2	2	1	2	1	2	2	2	0	1	2	2	1
Validity	17	15	12	1	7	0	4	10	9	8	7	9	8	14
Clarity	4	4	1	1	0	1	0	4	1	0	3	2	0	0
Scheduled review	2	2	0	1	1	0	1	1	1	1	2	1	1	1
Multidisciplinary process	6	4	4	1	0	1	1	5	5	0	0	3	1	1
Dissemination	1	0	0	0	0	0	0	0	1	0	0	0	0	0
Implementation	2	0	0	0	1	0	1	2	1	1	1	1	0	0
Evaluation	1	0	0	1	0	1	0	1	1	0	0	1	1	0
Other	1	1	0	0	0	0	0	0	0	0	0	1	0	0
Total no. of items	44	34	23	7	14	6	9	30	27	14	17	23	17	23

they differed considerably in terms of country of origin, rationale for development, intended purpose, number of questions/statements included in the instrument, and the particular aspects of the guideline they assessed (in terms of both items and guideline attributes).

From a methodologic perspective, there does not appear to be much evidence supporting the inclusion of most of the questions/statements in the instruments. Direct empirical support for inclusion of particular questions/statements was not given in the documentation provided with the instruments. While the rationale for including particular instrument questions/statements may be self-evident or based on recommendations of prominent organizations such as the IOM, more research showing that these questions/statements are objective or valid indicators of the guidelines' quality, or at the very least influence instrument users' perceptions of the guidelines' quality or utility, would be desirable. Nor does there appear to be much evidence of the reliability of the individual questions/statements. Only the instruments by Cluzeau et al. and Shaneyfelt et al. have been subjected to a validation study and shown to have some suggestion of validity (4;18;26). The Cluzeau instrument has also been shown to have good reliability.

The comparison also revealed that the items comprising the instruments largely focused on the process and format of guideline development, while fewer items assessed the overall clinical content or, more importantly, the clinical value of the guideline. Even guidelines that were properly developed by diligent and credible developers may not be clinically useful or as useful as they were expected to be. As has been noted by Cook and Giacomini (6), the usefulness of guidelines can only be determined with pilot testing or the dissemination and use of the guideline in clinical practice. The extent to which this type of information influences potential adopters' decisions to implement a guideline requires study.

Other items one would intuitively have expected to have been more prominent were the stated rationale or reason for the guideline and the specific purpose or aim of the guideline. Understanding why a guideline was developed and what it is expected to accomplish should help clinicians decide on its utility.

Another rarely occurring appraisal item that might also be expected to be important to clinicians and policy makers relates to the feasibility of implementing the guideline. Provision of such information might help overcome perceived clinical and organizational barriers that may influence clinicians' decisions to adopt a guideline. The importance of endorsement of guidelines by professional bodies is also underestimated by the existing instruments. The work of Tunis and colleagues (32) and Hayward and colleagues (11) suggests that the perceived credibility of guidelines is greatly influenced by who developed and endorsed them. Guidelines endorsed and/or developed by specialty bodies to which a potential guideline adopter belongs or which the adopter respects are considered more credible than guidelines developed by more general professional bodies, government, or industry.

This study has several limitations that must be acknowledged. The first relates to the generation of the items used in the content analysis of each instrument. We began by aggregating all the questions/statements from each instrument and then collapsing them into common items. This means that the results of the content analysis reflect how each instrument compares to all the others and not to a gold standard. An alternative approach might have been to use a panel of experts to come up with the ideal items that should be included in all guideline appraisal instruments. We chose the former approach since the generation of items for the content analysis by experts would likely not have been evidence-based, given the lack of evidence in this area, and would have probably required some form of consensus development. We did not have the resources to undertake such a consensus conference. The extent to which this procedure would have provided novel information is also questionable, since several instruments were originally developed using expert panels or committees (e.g., 17;20;26). Furthermore, our generation of items and the

subsequent assessment of each instrument in terms of these items was highly dependent on the explicitness of the wording of the questions/statements comprising each instrument. It was not uncommon that the same question/statement addressed two or more different items. For example, we interpreted the following question from the instrument of Hayward et al. (12), "Evidence: how and when evidence was gathered, selected and synthesized" as addressing the items for evidence collection, sources of evidence, literature selection, and data synthesis. For this reason, instruments that provided information on how to interpret each of their questions/statements may have fared better on the content analysis than instruments that did not provide documentation on how to interpret their questions/statements. The Cluzeau instrument (3) had an accompanying user guide that described how questions are to be interpreted.

We also did not evaluate the user friendliness or ease of use of the instruments. As noted above, there were differences in the understandability of instrument questions/statements but we did not explicitly evaluate the extent of these differences.

It must be remembered that assessing the quality of reporting of the guideline development process (which is what guideline appraisal instruments are in part focused on) may not be assessing the actual quality of this process. What guideline developers say about the development of a guideline may not do justice to how it was actually developed. Indeed, Cluzeau and colleagues (4;5) have noted that information about the process by which guidelines are developed is often left to a background document that may or may not be widely available.

CONCLUSIONS

We identified all guideline appraisal instruments retrievable by our methods and compared them for their ability to aid potential users to decide which ones to employ. Each of the instruments we considered had differing strengths and weaknesses in terms of its comprehensiveness in addressing all the guideline dimensions, completeness in the number of items addressing each attribute, and validity. It would seem that Cluzeau's instrument is the most well developed to date, and it has data to suggest that it is reliable and valid. It covered all 10 guideline attributes and addressed 28 of the 44 items generated by the content analysis. This instrument has been subject to validation studies and is currently being used by the National Health Service (NHS) Executive in the United Kingdom to help decide which guidelines to recommend to the NHS (4). The instrument also forms the basis of the Appraisal of Guidelines, Research and Evaluation in Europe (AGREE) instrument and will be evaluated in 10 European countries and Canada (19). Based on the results of this ongoing work, the instrument will likely be further modified and refined in the future. The Shaneyfelt instrument would appear to be the next well-developed and tested guideline appraisal instrument (26).

However, notwithstanding the considerable work that has gone into the development of the Cluzeau and Shaneyfelt instruments, the range of existing guideline appraisal instruments and the limited validation work that has been conducted to date, it is too early to confidently recommend the use of only one appraisal instrument. As it appears the role of clinical practice guidelines will continue to expand, those interested in appraising guidelines should select the instrument that best suits their situation and needs. The field of guideline appraisal would be greatly advanced by more research focusing on the validity and reliability of the various instruments and by experimental studies that compare, in a head-to-head fashion, the usefulness and outcomes of the different instruments.

Guidelines are developed to reduce practice variation and to improve clinical practice and patient care. Research needs to link the degree of uptake of guidelines with the nature of the guideline itself, in order that developers can learn how to create guidelines that can

be embraced by their targeted provider populations. Understanding the key attributes and items that a guideline should cover is a first step in this process. Looking ahead to the 21st century, professional and specialty bodies will have an opportunity to take a leadership role in not only developing but also evaluating practice guidelines developed by others. This role could include developing formal mechanisms for systematically reviewing and endorsing guidelines in their area of care. Such bodies, which tend to be national in scope, have the resources as well as the access to both the content and the methodologic expertise necessary to efficiently assess the quality and clinical utility of guidelines. They also have the necessary networks to disseminate their findings. By providing an "official guideline stamp of approval," professional bodies could help clinicians and policy makers sort out the value of specific guidelines and reduce the confusion resulting from conflicting guidelines.

REFERENCES

1. Calder L, Hébert P, Carter A, Graham I. Review of published recommendations and guidelines for the transfusion of allogeneic red blood cells and plasma. *Can Med Assoc J.* 1997;156:S1-S8.
2. Canadian Medical Association. *Guidelines for Canadian clinical practice guidelines*. Ottawa: Canadian Medical Association; 1994.
3. Cluzeau F, Littlejohns P, Grimshaw J, et al. Appraisal instrument for clinical guidelines. St. George's Hospital Medical School. Available at: http://sghms.ac.uk/phs/hceu/.
4. Cluzeau F, Littlejohns P, Grimshaw J, et al. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care.* 1999;11:21-28.
5. Cluzeau F, Littlejohns P, Grimshaw J, Hopkins A. Appraising clinical practice guidelines and the development of criteria: A pilot study. *J Interprofessional Care.* 1995;9:227-235.
6. Cook D, Giacomini M. The trials and tribulations of clinical practice guidelines. *JAMA.* 1999; 281:1950-1951.
7. Fahey T, Peters T. What constitutes controlled hypertension? Patient based comparisons of hypertension guidelines. *BMJ.* 1996;313:93-96.
8. Graham I, Beardal S, Tetroe J, Carter A, Laupacis A. The state of the art of practice guidelines development, dissemination, and evaluation in Canada [abstract]. Toronto: Scientific Basis for Health Services; 1997.
9. Grimshaw J, Freemantle N, Wallace S, et al. Developing and implementing clinical practice guidelines. *Qual Health Care.* 1995;4:55-64.
10. Grimshaw J, Russell I. Effect of clinical guidelines on medical practice: A systemic review of rigorous evaluations. *Lancet*. 1993;242:1317-1322.
11. Hayward RSA, Guyatt G, Moore K, et al. Canadian physicians' attitudes about preferences regarding clinical practice guidelines. *Can Med Assoc J.* 1997;56:1715-1723.
12. Hayward RSA, Wilson MC, Tunis SR, et al. More informative abstracts of articles describing clinical practice guidelines. *Ann Intern Med.* 1993;118:731-737.
13. Hayward RSA, Wilson MC, Tunis SR, et al. Users' guides to the medical literature, VIII: How to use clinical practice guidelines, A: Are the recommendations valid? *JAMA.* 1995;274:570-574.
14. Helwig A, Bower D, Wolff M, Guse C. Residents find clinical practice guidelines valuable as educational and clinical tools. *Fam Med.* 1998;30:431-435.
15. Institute of Medicine. *Clinical practice guidelines: Directions for a new program.* Washington, DC: National Academy Press; 1990.
16. Institute of Medicine. *Guidelines for clinical practice: From development to use.* Committee on Clinical Practice Guidelines. Washington, DC: National Academy Press; 1992.
17. Liddle J, Williamson M, Irwig L. Method for evaluating research guidelines evidence (MERGE). Sydney, Australia: New South Wales Department of Health; 1996. Available at http://www.health.nsw.gov.au/public health.
18. Littlejohns P, Cluzeau F, Bales R, et al. The quality and quantity of clinical practice guidelines for the management of depression in primary care in the UK. *Br J Gen Pract.* 1999;49:205-210.
19. Littlejohns P, Cluzeau F, Grimshaw J, et al. Promoting the rigorous development of clinical guidelines in Europe through the creation of a common appraisal instrument [abstract]. Amsterdam: Scientific Basis of Health Services; 1997.

20. Lohr K, Field M, eds. A provisional instrument for assessing clinical practice guidelines. In *Guidelines for clinical practice*. Washington, DC: National Academy Press; 1992.
21. Mendelson EB. The development and meaning of appropriateness guidelines. *Radiol Clin North Am.* 1995;33:1081-1084.
22. Perfetto EM, Morris LS. Agency for Health Care Policy and Research clinical practice guidelines. *Ann Pharmacother.* 1996;30:1117-1121.
23. Savoie I, Green CJ, Bassett K, et al. Critical appraisal criteria for clinical practice guidelines. Vancouver: BC Office of Technology Assessment; 1996.
24. Scottish Intercollegiate Guidelines Network (SIGN). *Clinical guidelines: Criteria for appraisal for national use.* Edinburg: Royal College of Physicians; 1995.
25. Selker HP. Criteria for adoption in practice of medical practice guidelines. *Am J Cardiol.* 1993;71:339-341.
26. Shaneyfelt T, Mayo-Smith M, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA.* 1999;281:1900-1905.
27. Steinhoff M, Abd El Khalek M, Khallaf N, et al. Effectiveness of clinical practice guidelines for the presumptive treatment of streptococcal pharyngitis in Egyptian children. *Lancet.* 1997;350:918-921.
28. Sudlow M, Thomson R. Clinical guidelines: Quantity without quality. *Qual Health Care.* 1997; 6:60-61.
29. Swales J. Guidelines on guidelines. *J Hypertens.* 1993;11:899-903.
30. Thomson R, Lavender M, Madhok R. How to ensure that guidelines are effective. *BMJ.* 1995; 311:237-342.
31. Thomson R, McElroy H, Sudlow M. Guidelines on anticoagulant treatment in atrial fibrillation in Great Britain: Variation in content and implications for treatment. *BMJ.* 1998;316:509-513.
32. Tunis SR, Hayward RSA, Wilson MC, et al. Internists' attitudes about clinical practice guidelines. *Ann Intern Med.* 1994;120:956-963.
33. Varonen H, Makela M. Practice guidelines in Finland: Availability and quality. *Qual Health Care.* 1997;6:75-79.
34. Ward J, Grieco V. Why we need guidelines for guidelines: A study of the quality of clinical practice guidelines in Australia. *Med J Aust.* 1996;165:574-576.
35. Woolf SH. Practice guidelines: What the family physician should know. *Am Fam Physician.* 1995;51:1455-1463.
36. Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: A systematic review. *Can Med Assoc J.* 1997;156:1705-1712.