

Is the methodological quality of guidelines declining in the US? Comparison of the quality of US Agency for Health Care Policy and Research (AHCPR) guidelines with those published subsequently

R Hasenfeld, P G Shekelle

Qual Saf Health Care 2003;12:428-434

Objective: To determine whether North American guidelines published subsequent to and in the same topic areas as those developed by the US Agency for Health Care Policy and Research (AHCPR) meet the same methodological criteria.

Study design: A guideline appraisal instrument containing 30 criteria was used to evaluate the methodological quality of the AHCPR guidelines, "updates" of the AHCPR guidelines authored by others, and guidelines that referenced or were adapted from the AHCPR guidelines. The frequency with which the criteria appeared in each guideline was compared and an analysis was performed to determine guidelines with two key features of the AHCPR guidelines—multidisciplinary guideline development panels and systematic reviews of the literature. Data were extracted from the guidelines by one investigator and then checked for accuracy by the other.

Results: Fifty two guidelines identified by broad based searches were evaluated. 50% of the criteria were present in every AHCPR guideline. The AHCPR guidelines scored 80% or more on 24 of the 30 criteria compared with 14 for the "updates" and 11 for those that referenced/adapted the AHCPR guidelines. All of the 17 AHCPR guidelines had both multidisciplinary development panels and systematic reviews of the literature compared with five from the other two categories ($p < 0.05$).

Conclusions: North American guidelines developed subsequent to and in the same topic areas as the AHCPR guidelines are of substantially worse methodological quality and ignore key features important to guideline development. This finding contrasts with previously published conclusions that guideline methodological quality is improving over time.

See end of article for authors' affiliations

Correspondence to:
R Hasenfeld, RAND, 1700
Main Street, P O Box
2138, Santa Monica, CA
90407-2138, USA;
rena@rand.org

Accepted for publication
4 July 2003

Although the development of clinical practice guidelines has exploded over the last 15 years, the methodological quality of such guidelines has recently attracted considerable attention.¹⁻³ A number of studies have examined the methods used to develop guidelines and found the majority to be inadequate. For instance, in a study of 279 guidelines identified in the peer reviewed medical literature from 1985 to 1997, Shaneyfelt *et al*¹ found that the mean overall adherence to standards by each guideline was only 43%, but that the methodological quality of guideline development was improving over time. Likewise, Grilli *et al*³ found that, of 431 guidelines developed by specialty societies from 1988 to 1998, only 33% reported information on the stakeholders involved in developing the guideline, 12% provided information on how the published evidence was identified, and 18% explicitly graded the strength of the recommendations. Only 5% of the guidelines in the study met all of these criteria. However, like Shaneyfelt, they also reported that the methods used to develop guidelines were improving over time. Similarly, Cluzeau *et al*² used an appraisal instrument to determine whether 60 guidelines published between 1991 and 1996 met specific methodological criteria and obtained respective mean scores of 34, 46, and 29 out of a possible total of 100 when testing for three dimensions of methodological quality.

Despite these poor findings, some guidelines have adhered to stricter methodological standards. It has generally been acknowledged that the 17 guidelines developed by the US

Agency for Health Care Policy and Research (AHCPR, now renamed the Agency for Healthcare Research and Quality (AHRQ)) between 1990 and 1996 used methods that significantly increased the quality of guideline development. However, in 1996 the AHCPR ended its guideline development program, a policy that the AHRQ has continued. Given this, an important question to consider is whether the higher standards that the AHCPR instituted will continue to be met without the ongoing presence of an authoritative body such as the AHRQ. We therefore undertook a study to examine those guidelines that have been published subsequent to and that deal with the same respective topic areas as the AHCPR guidelines to determine whether they meet these same methodological standards. Examining guidelines that have been published since the AHCPR guidelines is especially important since some of these later publications have specifically termed themselves or otherwise indicated that they are "updates" of the AHCPR guidelines, potentially leading users to assume that they are of the same methodological quality as the original AHCPR guidelines when indeed they may not be. To conduct our study we used a clinical practice guideline appraisal instrument to test whether the following types of North American guidelines met specific criteria indicative of good guideline development: (1) the 17 original AHCPR guidelines, (2) those guidelines that were authored by others but termed themselves "updates" of the AHCPR guidelines, and (3) any guidelines that referenced or were adapted from the AHCPR guidelines.

METHODS

Identification and development of review criteria

We chose to use the clinical practice guideline appraisal instrument developed by Shaneyfelt *et al*³ which was deemed to be one of the two best developed instruments in a comparative assessment by Graham *et al*.⁴ The AGREE instrument incorporates many of the same concepts as the Shaneyfelt instrument with respect to scope and purpose, stakeholder involvement, rigor of development, and clarity and presentation.³ In addition, the AGREE instrument includes items regarding the application of the guidelines (such as pilot testing and tools for application). We added to the Shaneyfelt instrument the following additional criteria that we considered important factors in the guideline development process:

- Are the participants (in the guideline development process) multidisciplinary?
- Is there a definition for the strength of evidence (the recommendations) use?
- Was there a systematic review of the evidence?

The first and third of these criteria correspond to AGREE items 4 (the guideline development group includes individuals from all the relevant professional groups) and 8 (systematic methods were used to search for evidence), respectively; we determined that the second criterion was necessary since empirical evidence has shown that using different “strength of evidence” systems can result in different recommendations⁶ so users need to know the definition that was used. In addition, the Shaneyfelt instrument includes two criteria regarding the benefits and harms of the health practices specified in a guideline and whether or not these are quantified; we retained these criteria but listed benefits separately—for example, benefits of specific health practices are specified, harms of specific health practices are specified. The guideline criteria we used for the study are listed in table 1.

Identification of guidelines

For each of the AHCPR guideline topics (Alzheimer’s disease, benign prostatic hyperplasia, cardiac rehabilitation, cataracts, depression, heart failure, low back pain, mammography, otitis media with effusion, pain (acute and cancer), post-stroke rehabilitation, pressure ulcers (treatment and prevention), sickle cell anemia, unstable angina, and urinary incontinence), in 2001/2 we conducted literature searches in MEDLINE and Embase from the year in which each respective AHCPR guideline was published up to April 2000, limiting these searches to articles designated as “practice guidelines” or “guidelines”. Since clinical practice guidelines often are not referenced in the traditional scientific databases, we used a number of additional search strategies. Primarily, we used the National Guideline Clearinghouse on the Internet to identify guidelines that fit our criteria. When possible, we downloaded the appropriate guidelines from the National Guideline Clearinghouse or the publishing organizations’ websites. When this was not possible we reviewed the summary statements provided by the National Guideline Clearinghouse for each guideline. In addition, we searched the following sources: Computerized Needs-Oriented Quality Measurement Evaluation System (CONQUEST) 1.1, A Project to Develop and Evaluate Methods to Promote Ambulatory Care Quality (DEMPAQ), Directory of Clinical Practice Guidelines, Guide to Clinical and Preventive Services, Health Plan Employer Data and Information Set (HEDIS) 3.0, National Library of Health Care Indicators, and the Medical Outcomes and Guidelines Sourcebook.^{7–13} We also requested guidelines from the following agencies and

organizations: Administration on Aging, Centers for Disease Control and Prevention, Department of Health and Human Services, Department of Veterans’ Affairs, Foundation for Accountability, Health Care Financing Administration (HCFA), HCFA/Connecticut Peer Review Organization, Joint Commission on Accreditation of Healthcare Organizations, National Committee for Quality Assurance, and National Institutes of Health. Finally, we identified guidelines from professional societies, disease societies, medical organizations, medical associations, and patient advocacy groups. When possible, we located these groups on the Internet and downloaded the guidelines; otherwise, we telephoned or wrote to them to request the materials.

Guideline inclusion/exclusion criteria

All of the 17 most recently published AHCPR guidelines were included. We excluded the recently published document “*Treating tobacco use and dependence: a clinical practice guideline*” because this is a Public Health Service guideline and not an AHCPR guideline. We also excluded any publications that were written outside North America because the AHCPR is an American government agency and we assumed that its influence was likely to be greater on publications originating in the US and Canada.

We selected additional guidelines for the study in three phases. First, although we included the terms “practice guidelines” or “guidelines” in our literature search, it was necessary to clearly identify those documents that met our definition of a “practice guideline”—namely, any document that identified itself as such or contained specific recommendations for healthcare decisions that were not authored by one or more people writing as individuals (to distinguish the document from a review article). Next we included guidelines that were published after the AHCPR documents since our research question was to assess whether later guidelines met the same methodological standards as the AHCPR guidelines. Of these guidelines, we chose those that were authored by others but termed themselves as “updates” of the AHCPR guidelines or those that referenced or were adapted from the AHCPR guidelines since readers or users of these guidelines are most likely to infer that some of the rigor of the original AHCPR guidelines carried over to the new guidelines. We included the guidelines themselves and any supporting material available with it but did not consider supplemental materials that required accessing a different source.

Scoring the guidelines

Once we had identified the guidelines we coded them as (1) the original AHCPR guidelines, (2) guidelines that termed themselves as updates of the AHCPR guidelines, and (3) guidelines that referenced or adapted the AHCPR guidelines. We then reviewed each guideline to determine whether it contained the criteria in our guideline appraisal instrument, with one researcher conducting the initial review and the second researcher verifying the results. For each guideline we created a chart listing the criteria. When the criteria were present in the guideline we marked the appropriate boxes on the chart and indicated the page number(s) on which we found the information.

In order to statistically compare a summary score between the guidelines we created a binary “yes/no” outcome for each guideline depending on whether it met the following two criteria that we and others have judged as fundamental methodological attributes that distinguish the AHCPR guidelines:

- a systematic review of the evidence (criterion 23a in table 1); and

Table 1 Percentage (number) of guidelines containing the guideline appraisal instrument criteria by category

Guideline appraisal instrument criteria	All guidelines (N = 52)	AHCPR guidelines (N = 17)	Updates of AHCPR guidelines (N = 5)	Guidelines that reference/adapt AHCPR guidelines (N = 30)
(1) Purpose of the guideline is specified	100 (52)	100 (17)	100 (5)	100 (30)
(2) Rationale and importance of the guideline are explained	77 (40)	100 (17)	100 (5)	60 (18)
(3) Participants in the guideline development process and their areas of expertise are specified	96 (50)	100 (17)	80 (4)	97 (29)
(3a) Are the participants multidisciplinary?*	73 (38)	100 (17)	40 (2)	63 (19)
(4) Targeted health problem or technology is clearly defined	100 (52)	100 (17)	100 (5)	100 (30)
(5) Targeted patient population is specified	100 (52)	100 (17)	100 (5)	100 (30)
(6) Principal preventive, diagnostic, or therapeutic options available to clinicians and patients are specified	100 (51)**	100 (16)**	100 (5)	100 (30)
(7) Method by which the guideline underwent external review is specified	58 (30)	100 (17)	60 (3)	33 (10)
(8) Expiration date or date of scheduled review is specified	29 (15)	12 (2)	20 (1)	40 (12)
(9) Method for grading or classifying the scientific evidence is specified	46 (24)	59 (10)	40 (2)	40 (12)
(10a) Benefits of specific health practices are specified	100 (52)	100 (17)	100 (5)	100 (30)
(10b) Harms of specific health practices are specified	98 (51)	100 (17)	80 (4)	100 (30)
(11a) Benefits are quantified	64 (33)	94 (16)	100 (5)	40 (12)
(11b) Harms are quantified	64 (33)	94 (16)	80 (4)	43 (13)
(12) Recommendations are graded according to the strength of evidence	46 (24)	82 (14)	40 (2)	27 (8)
(12a) Is there a definition for the strength of evidence they use?*	46 (24)	82 (14)	40 (2)	27 (8)
(13) Effect on health care costs from specific health practices is specified	44 (23)	94 (16)	20 (1)	20 (6)
(14) Costs are quantified	37 (19)	82 (14)	20 (1)	13 (4)
(15) Role of patient preferences is discussed	60 (31)	82 (14)	40 (2)	50 (15)
(16) Health outcomes are specified	100 (52)	100 (17)	100 (5)	100 (30)
(17) Flexibility in the recommendations is specified	87 (45)	100 (17)	40 (2)	87 (26)
(18) Evidence used is identified by citation and referenced	92 (48)	100 (17)	100 (5)	87 (26)
(19) Formal methods of combining evidence or expert opinion are used and described	2 (1)	0 (0)	0 (0)	3 (1)
(20) Recommendations are specific and apply to the stated goals of the guideline	98 (51)	94 (16)	100 (5)	100 (30)
(21) Role of value judgements used by the guideline developers in making recommendations is discussed	14 (7)	6 (1)	0 (0)	20 (6)
(22) Intended audience or users of the guideline are specified	83 (43)	94 (16)	80 (4)	77 (23)
(23) Method of identifying evidence is specified	54 (28)	100 (17)	60 (3)	27 (8)
(23a) Was there a systematic review of the evidence?*	52 (27)	100 (17)	60 (3)	23 (7)
(24) Time period from which evidence is reviewed is specified	39 (20)	47 (8)	60 (3)	30 (9)
(25) Method of data extraction is specified	21 (11)	47 (8)	20 (1)	7 (2)

*We added these criteria to the clinical practice guideline appraisal instrument developed by Shaneyfelt *et al.*

**Since we determined that this criterion was not applicable to the AHCPR guideline on mammography, the total number of guidelines for this criterion was 51 for all of the guidelines and 16 for the AHCPR guidelines.

Table 2 Number of guidelines included in the study by category and condition

Condition	Total no of guidelines	AHCPR guidelines	Updates of AHCPR guidelines	Guidelines that reference/adapt AHCPR guidelines
Alzheimer's disease	4	1	0	3
Benign prostatic hyperplasia	1	1	0	0
Cardiac rehabilitation	2	1	0	1
Cataracts	3	1	0	2
Depression	5	1	1	3
Heart failure	7	1	2	4
Low back pain	2	1	0	1
Mammography	1	1	0	0
Otitis media with effusion	2	1	0	1
Pain (acute and cancer)	7	2	0	5
Post-stroke rehabilitation	1	1	0	0
Pressure ulcers (prevention and treatment)	7	2	1	4
Sickle cell anemia	2	1	0	1
Unstable angina	4	1	1	2
Urinary incontinence	4	1	0	3
Total	52	17	5	30

- the use of a multidisciplinary group of experts to interpret the evidence (criterion 3a in table 1). For a guideline panel to be defined as multidisciplinary we required only that at least one specialist and one generalist participated in the guideline development process. We defined a generalist as any physician who practiced internal medicine, pediatrics, or obstetrics/gynecology. If this information was not specified in the guideline, we conducted Internet searches on the participants to determine their practice areas.

As a sensitivity analysis, we also included in our comparisons those guidelines that graded the strength of their recommendations (criterion 12 in table 1).

Data analysis

Univariate frequencies were first computed for each of the guideline criteria. We then performed cross tabulations, including χ^2 tests, comparing the frequency with which the criteria appeared in the AHCPR guidelines with the frequency with which the criteria appeared in each of the other guideline categories. In addition, analysis of variance and *t* tests were used to assess possible differences in the item scores grouped according to the following domains, as defined by Shaneyfelt *et al*¹:

- guideline development and format (criteria 1, 2, 3, 3a, 4, 5, 6, 7, 8, 16, and 22 in table 1);
- evidence identification and summary (criteria 9, 10a, 10b, 11a, 11b, 13, 14, 18, 19, 23, 23a, 24, and 25 in table 1); and
- formulation of the recommendations (criteria 12, 12a, 15, 17, 20, and 21 in table 1).

We also conducted a separate analysis comparing the guidelines based on whether they had both multidisciplinary guideline development panels and systematic reviews of the literature. To do this we determined those guidelines that met both of these criteria and created a table that compared them by guideline category. We then conducted a χ^2 test of the entire table to compare overall differences and *t* tests to compare differences between the three guideline categories.

RESULTS

Guidelines identified

As fig 1 indicates, our MEDLINE and Embase searches yielded a total of 6200 documents and our alternative search

strategies (such as the National Guideline Clearinghouse, the Internet, medical associations, professional societies, etc) identified a further 159, giving a total of 6359 publications (including both articles and guidelines) that dealt with the same topic areas as and were published after the AHCPR guidelines. From these documents we identified 86 that were updates of or referenced/adapted the AHCPR guidelines. Of this subset, 35 documents fit our definition of a clinical practice guideline and were not published outside North America. Thus, including the 17 AHCPR guidelines, we identified 52 guidelines that met all of our inclusion criteria. Table 2 shows the number of guidelines included in the study by health condition and guideline category. Most of the documents in the study (58%) were those that referenced or were adapted from the AHCPR guidelines; only five (10%) were updates of the AHCPR guidelines. A comprehensive listing of the guidelines included in the study by category can be found in Appendix 1 on the QSHC website (www.qshc.com/supplemental).

Guideline criteria

Table 1 shows the number of guidelines that met the criteria in our guideline appraisal instrument by guideline category. (A list of all the guidelines by category and the specific criteria that each guideline met are shown in Appendix 2 which is available on the QSHC website at www.qshc.com/supplemental.) None of the guidelines fulfilled all of the criteria, but a number of criteria were present in most of the guidelines that we evaluated.

The following six criteria appeared in all of the guidelines:

- a statement of purpose (criterion 1);
- a definition of the health problem or technology addressed in the guideline (criterion 4);
- a description of the targeted patient population (criterion 5);
- a specification of the principal preventive, diagnostic, or therapeutic options available (criterion 6; it should be noted that we determined that this criterion was not applicable to the AHCPR mammography guideline, as indicated in table 1);
- a description of the benefits from the health practices specified in the guideline (criterion 10a); and
- an explanation of the expected health outcomes from the recommendations in the guideline (criterion 16).

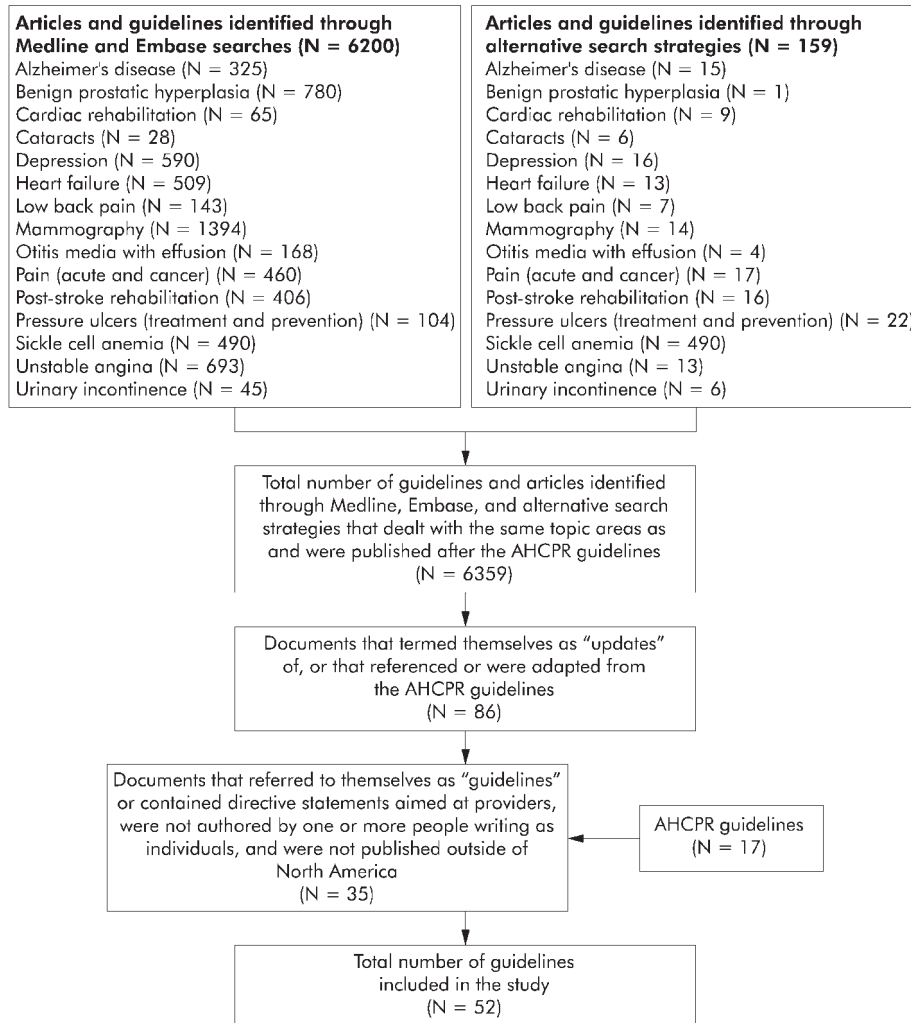


Figure 1 Guidelines identified for study.

Fifty one of the guidelines (98%) specified the harms that could occur from the recommendations in the guideline (criterion 10b) and included recommendations that were specific and applied to the guideline’s stated goals (criterion 20). Almost all of the guidelines (96%) listed the participants in the guideline development process and their areas of expertise (criterion 3), although only 73% of the participants were multidisciplinary (criterion 3a). In addition, 92% of the guidelines identified and referenced the evidence used to develop the guideline (criterion 18).

Although all of the criteria in the appraisal instrument were present in at least one of the guidelines in the study, several measures appeared infrequently. Only one of the guidelines that we reviewed used and described formal methods of combining evidence or expert opinion (criterion 19). In addition, very few of the guidelines (14%) discussed the role of the developers’ value judgements in making recommendations (criterion 21). Likewise, the method of data extraction used in preparing the guideline (criterion 25) was described in only 21% of the guidelines and an expiration date or a date of scheduled review (criterion 8) appeared in only 29%.

When we analyzed our results by guideline category we found that the AHCPR guidelines contained the largest number of criteria. 50% of the measures were present in all of the guidelines in this category compared with only 33% and 27% in guidelines that were updates and those that

referenced/adapted the AHCPR guidelines, respectively. This difference was particularly apparent for two measures pertaining to evidence: all of the AHCPR guidelines specified the method they used to identify evidence (criterion 23) and had conducted a systematic review of the literature (criterion 23a); in contrast 60% of the updates and only 27% and 23%, respectively, of the guidelines that referenced or adapted the AHCPR publications met these two criteria. In addition, all of the AHCPR guidelines had multidisciplinary panels (criterion 3a) and specified the process by which the guideline underwent external review (criterion 7), while only 40% and 60% of the updates and 63% and 33% of the guidelines that referenced or adapted the AHCPR documents met these respective criteria.

Overall, the AHCPR guidelines ranked highest on the appraisal instrument. The AHCPR guidelines scored 80% or more on 24 of the 30 criteria, while the updates and the guidelines that referenced or adapted the AHCPR documents scored 80% or more on 14 and 11 of the criteria, respectively. The AHCPR guidelines scored most poorly and were outscored by one or both of the other categories on three measures:

- none of the AHCPR documents used and described formal methods of combining evidence or expert opinion (criterion 19) while one (3%) guideline that referenced/adapted the AHCPR documents did so;

Key messages

- North American clinical practice guidelines published subsequent to and in the same respective topic areas as those published by the US Agency for Health Care Policy and Research (AHCPR) are of poorer methodological quality, even when they term themselves as “updates” or in some way infer that they are related to the original AHCPR guidelines. Thus, users may be misled into thinking that these later guidelines are of similar methodological quality as the AHCPR guidelines when this is not the case.
- Most newer guidelines did not conduct a systematic review of the literature or use multidisciplinary expert panels—two criteria that have been empirically linked to bias and appear in all of the AHCPR guidelines.
- Without the benefit of an oversight agency such as the AHCPR, it appears that the methodological quality of clinical practice guidelines will remain less than acceptable.

- only two (12%) of the ACHPR documents specified an expiration date or date of scheduled review (criterion 8) while one (20%) of the updates and 12 (40%) of the guidelines that referenced/adapted the ACHPR documents did so; and
- only one (6%) of the AHCPR documents discussed the role of value judgements used by the developers in making recommendations (criterion 21) while six (20%) of the guidelines that referenced/adapted the AHCPR documents did so.

In addition, our analysis of the criteria grouped according to the domains defined by Shaneyfelt *et al* showed that the AHCPR guidelines scored highest in all three domains. For the first two domains (guideline development and format, and evidence identification and summary) the differences were statistically significant between the AHCPR guidelines and the guidelines that referenced or adapted them; for the third domain (formulation of the recommendations) the differences were statistically significant between the AHCPR guidelines and both of the other guideline categories.

The AHCPR guidelines also outranked the other guideline categories when we narrowed our comparisons to two criteria—whether the guideline development panels were multidisciplinary (criterion 3a) and whether a systematic review of the evidence had been conducted (criterion 23a). The results of these comparisons are shown in table 3. All of the AHCPR documents met both of these criteria compared with only one (20%) of the updates and four (13%) of the guidelines that referenced or were adapted from the AHCPR

Table 3 Percentage (number) of guidelines developed using systematic reviews and multidisciplinary panels

	AHCPR guidelines (N = 17)	Updates of AHCPR guidelines (N = 5)	Reference/adapt AHCPR guidelines (N = 30)
Yes	100 (17)	20 (1)	13 (4)
No	0 (0)	80 (4)	87 (26)

p<0.05 for all.
Yes = systematic review of the literature and multidisciplinary participants in the guideline development process.

documents. All of these comparisons were statistically significant (p<0.05). The AHCPR guidelines also did considerably better than the other categories when we added grading the strength of the recommendation as a third criterion for a sensitivity analysis. Fourteen of the AHCPR guidelines (82%) met all three of these criteria compared with one (20%) of the updates and two (7%) of the guidelines that referenced or adapted the AHCPR guidelines (p<0.05). Finally, when we restricted the comparison solely to guidelines that had conducted a systematic review of the literature we found that all the AHCPR guidelines met this criterion compared with 60% of the updates (p = 0.006) and 23% of the guidelines that referenced or adapted the AHCPR guidelines (p<0.0001).

DISCUSSION

In contrast to the findings of Shaneyfelt *et al*¹ and Grilli *et al*³ that the methodological quality of guidelines has been improving over time, we found that newer guidelines in the same topic areas as the AHCPR guidelines were sharply and disturbingly poorer in methodological quality than the older AHCPR guidelines. When we used a clinical practice guideline appraisal instrument to evaluate 52 guidelines published subsequent to and in the same topic areas as the AHCPR guidelines, this finding held true for guidelines that called themselves “updates” of the AHCPR guidelines, as well as for those that either referenced or were adapted from the AHCPR guidelines. Thus, guidelines that call themselves “updates” or in some other way infer that they are related to the original AHCPR guidelines mislead readers into thinking that they are of similar methodological quality when this is not the case.

The lack of adherence to AHCPR methodological standards that we observed among more recently published guidelines may have important consequences. Patients cannot reap the full benefits from health care that is delivered according to clinical practice guidelines when those guidelines do not reflect the most accurate, up to date, and unbiased synthesis of the scientific literature and expert opinion. To the extent that the decrease in the quality of guideline development has led to more biased guidelines, it is more likely that patient care and outcomes will suffer. Two of the criteria in our appraisal instrument— systematic reviews of the literature and multidisciplinary expert panels—have empirical evidence associating them with bias, and most of the newer guidelines in our study did not meet these criteria.^{14–16}

The groups developing the new or “updated” guidelines probably did not have the same level of resources available to them as the AHCPR. Whether equally valid guidelines can be developed by skipping some resource intensive steps has not been conclusively evaluated. However, it has been shown that some resource intensive processes such as a systematic as opposed to a traditional literature review do reach different conclusions, and the sanguine acceptance of skipping such steps due to resource constraints cannot, in our opinion, be supported.

Our study has several limitations that are worth noting. Firstly, although we used a number of methods to identify guidelines that were published subsequent to and in the same respective topic areas as the AHCPR guidelines, there is the possibility that we may have missed or overlooked some. However, if the guidelines were not published in a major journal or readily available through the National Guideline Clearinghouse, then most potential users would probably miss them as well. Secondly, we relied only on the guidelines themselves and any supporting materials provided with the guidelines when assessing quality. We did not assess supplementary materials that may have been published elsewhere. Thirdly, the methodological criteria on which we evaluated the guidelines are based on expert opinion and few

have been shown to be predictive of validity. However, our guideline appraisal instrument included two criteria for which there is empirical evidence of bias—multidisciplinary panels and systematic reviews of the literature—and our analysis restricted to these two criteria was consistent with our analysis using all of the criteria (the newer guidelines scored much worse than the AHCPR guidelines). Fourth, we had to rely on our own judgement to determine whether or not the guidelines in the study met the criteria in the appraisal instrument. We attempted to mitigate this subjectivity by having one researcher check for the criteria in the guidelines and the second researcher to verify the results. Finally, our use of the National Guideline Clearinghouse (NGC) to identify guidelines for our study may have biased our selections upwards since one of the NGC's selection criteria is a systematic review of the literature. Nevertheless, we used a broad range of sources to identify guidelines and we still found that only 60% of the updates and 23% of the guidelines that referenced or adapted the AHCPR guidelines had conducted a systematic review of the literature.

Despite these limitations, our study has several strengths. We reviewed all of the AHCPR guidelines and thus covered a wide range of health conditions. The only AHCPR guideline that we excluded from the study was the one dealing with HIV infection which we chose to omit because medical practice in this area changes so rapidly that no one guideline can remain up to date. In addition, we assessed the guidelines using criteria obtained from one of the most highly rated clinical practice guideline appraisal instruments.¹⁻⁴ Finally, our attempts to identify guidelines that met our study criteria were very broad and included extensive searches using a multitude of sources.

This study shows that newer guidelines pertaining to the same topics as the AHCPR guidelines do not meet the same methodological standards. This is the case even when the more recent guidelines have called themselves “updates” of the AHCPR guidelines. It therefore appears that the process of guideline development in these clinical areas has regressed substantially since the publication of the AHCPR guidelines. This decline does not serve patients or clinicians, both of whom increasingly rely upon clinical practice guidelines to obtain information about proper treatment. One possible solution to this problem is the formation of a central authority—akin to the evidence-based practice center program that promotes standards for literature synthesis—that would promote the development of guidelines that adhere to high methodological standards. Indeed, without such an authority it is difficult to see how the methodological quality of guidelines can be maintained. Having a single agreed method for appraising guidelines would be an appropriate starting place so that guidelines developed by different groups and/or in different countries can be compared. The AGREE instrument probably comes closest to an international standard at this time, although it is used less frequently in North America than it is in Europe. Increasing the rigor with which guidelines are developed comes at a cost, however, and unfortunately not every organization can bear the expenses involved in creating methodologically sound guidelines. If this is the case, we must consider whether patients, physicians, and policymakers are well served by guidelines that are developed without the proper resources to ensure their quality.

ACKNOWLEDGEMENTS

The authors acknowledge Elizabeth Roth and Sally Morton for providing the statistical analyses for the study, Roberta Shanman for conducting literature searches, Nadia Khatib for inputting the data, and Shannon Rhodes for providing project support. This article is based in part on work performed by the Southern California Evidence Based Health Center under contract to the Agency for Healthcare Research and Quality (contract no 290-97-0001), Rockville, MA, USA.



A comprehensive listing of the guidelines included in the study by category can be found in Appendix 1 and a list of all the guidelines by category and the specific criteria that each guideline met are shown in Appendix 2, both of which are available on the QSHC website at www.qshc.com/supplemental.

Authors' affiliations

R Hasenfeld, Southern California Evidence-Based Practice Center, RAND Health, Santa Monica, California, USA

P G Shekelle, Southern California Evidence-Based Practice Center, RAND Health, Santa Monica, California; Department of General Internal Medicine, Greater Los Angeles VA Medical Center, Los Angeles, California; Senior Research Associate, the Veterans Affairs Health Services Research and Development Service, USA

REFERENCES

- 1 **Shaneyfelt TM**, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999;**281**:1900–5.
- 2 **Cluzeau FC**, Littlejohns P, Grimshaw JM, *et al*. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999;**11**:21–8.
- 3 **Grilli R**, Magrini N, Penna A, *et al*. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000;**355**:103–6.
- 4 **Graham ID**, Calder LA, Hebert PC, *et al*. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care* 2000;**16**:1024–38.
- 5 **AGREE Collaboration**. *Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument* (available at www.agreecollaboration.org).
- 6 **Ferreira PH**, Ferreira ML, Maher CG, *et al*. Effect of applying different 'levels of evidence' criteria on conclusions of Cochrane reviews of interventions for low back pain. *J Clin Epidemiol* 2002;**55**:1126–9.
- 7 **Agency for Health Care Policy and Research**. *CONQUEST 1.1 (Computerized Needs-Oriented Quality Measurement Evaluation System)*. Agency for Health Care Policy and Research, Public Health Service, Department of Health and Human Services, 1996.
- 8 **Palmer RH**, Clark LE, Lawthers AG, *et al*. *DEMPAQ: a project to Develop and Evaluate Methods to Promote Ambulatory Care Quality*. Final Report. Volumes I, II, and III. Boston, MA: Harvard School of Public Health, 1994.
- 9 **Toepp MC**, Kuznets N, Herrera S. *Directory of clinical practice guidelines: titles, sources, and updates*. Chicago, IL: American Medical Association, 1998.
- 10 **US Preventive Services Task Force**. *Guide to clinical preventive services*, 2nd ed. Alexandria, VA: International Medical Publishing, 1996.
- 11 **Committee on Performance Measurement**. *HEDIS 3.0: Health Plan Employer Data and Information Set*. Washington, DC: National Committee for Quality Assurance, 1997.
- 12 **Joint Commission on Accreditation of Healthcare Organizations**. *National Library of Healthcare Indicators: Health Plan and Network Edition*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations, 1997.
- 13 **Anon**. *The medical outcomes and guidelines sourcebook: a progress report and resource guide on medical outcomes research and practice guidelines: developments, data, and documentation*. New York: Faulkner and Gray, 1997.
- 14 **Dickersin K**, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286–91.
- 15 **Kahan JP**, Park RE, Leape LL, *et al*. Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Med Care* 1996;**34**:512–23.
- 16 **Leape LL**, Park RE, Kahan JP, *et al*. Group judgments of appropriateness: the effect of panel composition. *Qual Assur Health Care* 1992;**4**:151–9.