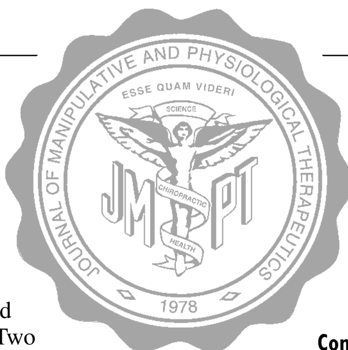
170 Journal of Manipulative and Physiological Therapeutics
Volume 24 • Number 3 • March/April 2001
0161-4754/2001/$35.00 + 0 **76/1/113775** © 2001 JMPT

Evaluating the Quality of Clinical Practice Guidelines

Jeffrey R. Cates, DC, MS,[a] David N. Young, DC, PhD,[b] David J. Guerriero, DC, MS,[c] Warren T. Jahn, DC, MPS,[d] Jesse P. Armine, RN, DC,[e] Alan B. Korbett, DC, DO,[f] Daniel S. Bowerman, DC,[g] Robert C. Porter, MD,[h] Terry D. Sandman, DC, MPH,[i] and Robert A. King, DC[j]



ABSTRACT

Objective: To review and identify established methods for evaluating the quality of practice guidelines and to use a selected assessment tool to assess 2 chiropractic practice guideline documents.

Methods: A search of the medical literature was performed to identify current methods and procedures for practice guideline evaluation. Two chiropractic practice guideline documents, *Vertebral Subluxation in Chiropractic Practice* (CCP) and *Guidelines for Chiropractic Quality Assurance and Practice Parameters* (Mercy) were then independently evaluated for validity by 10 appraisers using the identified appraisal tool. The appraisal scores were tabulated, and consensus appraisals were generated for the CCP and Mercy guideline documents.

Results: The "Appraisal Instrument for Clinical Guidelines" (Cluzeau instrument) was identified as a reliable and valid method of guideline evaluation. The result of the application of this appraisal tool in the assessment of the CCP and Mercy guideline documents was that the former scored notably lower than the latter. On the basis of the results of the guideline appraisals, the CCP document is not recommended, and its guidelines are not considered suitable for application in chiropractic practice. The Mercy guidelines are recommended for application in chiropractic practice, with the proviso that new scientific data should be considered.

Conclusions: The literature reviewed suggests that professional organizations or groups should undertake a critical review of guidelines using available critical guideline appraisal tools. Guideline validity appraisal should be done before acceptance by the chiropractic profession. To avoid unwarranted utilization of poorly constructed guidelines, it is strongly recommended that all future guidelines be reviewed for validity and scientific accuracy with the findings published in a medically indexed journal before they are adopted by the chiropractic community. (J Manipulative Physiol Ther 2001; 24:170-6)

Key Indexing Terms: Appraisal; Chiropractic; Clinical Practice; Guidelines; Reliability; Validity

INTRODUCTION

There has been a prodigious production of practice guidelines relating to health care issues in recent years. With so many sets of guidelines being produced by so many organizations, physicians now find that they need to assess guidelines' quality and validity before using them in practice. The question of how to evaluate clinical guidelines then arises. We will entertain the hypothesis that a method of evaluating guidelines exists by which chiropractic guideline quality can be reliably evaluated.

Practice guidelines are expert or evidence-based recommendations regarding optimal clinical protocols for a health care field or condition under treatment. Guidelines, which can be either procedure-based or condition-based, are considered valid when they optimize patient care and result in health gain at the expected costs. Numerous guideline documents regarding chiropractic practice have been developed; they have varying degrees of overall quality, validity, and clinical applicability.[1-3] High-quality practice guidelines clarify what interventions are effective in various circumstances and specify the extent of the documentation supporting those interventions. Guidelines can also reveal those procedures and interventions that are poorly supported by the scientific literature or require additional research to demonstrate efficacy. Flawed guidelines can result in ineffective or inappropriate treatment, overutilization, or poor clinical outcomes and can harm patients by disseminating information and advice that is scientifically inaccurate.[4] Health care guidelines should be subjected to peer review and testing by professional organizations or groups before

[a]Private Practice of Chiropractic Orthopedics, Oregon, Ill.
[b]Private Practice of Chiropractic Orthopedics, Chula Vista, Calif.
[c]Private Practice of Chiropractic Medicine, Orlando, Fla.
[d]Private Practice of Chiropractic Orthopedics and Sports, Roswell, Ga.
[e]Private Practice of Chiropractic, Havertown, Pa.
[f]Private Practice of Adult, Child and Adolescent Psychiatry, Snellville, Ga.
[g]Private Practice of Chiropractic Orthopedics, Philadelphia, Pa.
[h]Private Practice of Occupational Medicine, Rockford, Ill.
[i]Private Practice of Radiology, Roselle, Ill.
[j]Private Practice of Chiropractic and Quality Assurance and Utilization Review, Purcellville, Va.
Submit reprint requests to: Jeffrey R. Cates, DC, 200 N. 6th Street, Oregon, IL 61061. E-mail: *cates@essex1.com*.
Paper submitted February 7, 2000; in revised form March 6, 2000.
doi:10.1067/mmt.2001.113775

the guidelines' validation and application in clinical practice to maximize the benefit and limit any potential harm arising from the use of flawed guidelines.[5-7]

The development of clinical practice guidelines has evolved from basing recommendations on expert opinions to relying on a systematic review of evidence regarding the efficacy of various procedures.[8] Inasmuch as guidelines now rely more on documented evidence and less on opinions, it has become possible to evaluate the quality of a guideline in a method similar to meta-analysis.[9]

It is our intent to identify and evaluate some currently accepted guideline evaluation techniques and to apply them to 2 chiropractic guideline documents: *Vertebral Subluxation in Chiropractic* (CCP),[10] produced by the Council on Chiropractic Practice, and *Guidelines for Chiropractic Quality Assurance and Practice Parameters* (Mercy),[11] forged at the Mercy Center Consensus Conference. The purpose of our work is to (1) identify an effective evaluation instrument and (2) use the instrument to evaluate these 2 chiropractic guideline documents. It is hoped that this work will serve to identify a method that can be used to evaluate the quality of these and future chiropractic practice guidelines.

METHODS

Literature Review

A search of the medical literature was performed to identify current methods and applications of guideline evaluation. This included a computer search and a hand search for literature pertaining to guideline reviews. We found a significant amount of general information that was helpful in assessing guidelines and a few organized procedures and methods that were deemed applicable to our task.[1,3-8,11-32] A study by Graham et al[32] reviewed the current literature relating to guideline evaluation. They identified and indexed 13 guideline appraisal instruments from 5 different countries. This work is notable because it reviews the current status of guideline evaluation instruments, including validity and reliability issues. Graham et al[32] evaluated assessment instruments according to their ability to assess 10 major categories of desirable guideline attributes. The 10 attributes are based on the recommendations of the Institute of Medicine (IOM) regarding desirable guideline attributes.

The available literature indicates that guideline validity and applicability is dependent on 3 essential elements: the nature and composition of the panel formulating the guidelines, the validity of the clinical evidence used to formulate the guidelines, and the construction of the guideline document itself.

Guideline development panels should include experts in clinical content, systematic review, and guideline development as well as potential users of the guidelines.[26] Hayward et al[5] warned that panels dominated by single groups without diversification can result in intellectually and financially biased guidelines. In addition, Shekelle et al[26] recommended that to limit individual bias, the process should have input from all groups whose activities would be impacted by acceptance of the guidelines.

An opinion common in all of the literature reviewed is that practice guidelines should be based on solid scientific evidence.[1,3,12,22,23] Recommendations based solely on opinions, clinical judgment, and experience are those most susceptible to bias and self-interest.[5,26] To help guideline users evaluate whether a reasonable and rational method was used in gathering and interpreting the evidence, Hayward et al[5] recommended that a clear overview of the criteria used for selecting evidence be included in any set of guidelines. Clear and logical boundaries for determining the level of admissible evidence and the strength of that evidence must exist.[5,16,26] The process must include the removal of irrelevant articles and the identification of papers with flawed methods.[16] Eccles et al[16] recommended a review of evidence with an emphasis on meta-analysis, randomized controlled trials, reviews, and both controlled cohort studies and controlled case studies.

Hayward et al[5] recommended that guidelines should report the degree of consensus reached by the panel as well as strategies for reporting dissenting or minority opinions. Shekelle et al[26] recommended external review of guidelines, and others have recommended actual testing of guidelines in clinical practice.[22,25]

Some of the more intriguing guideline evaluation methods identified include the IOM's "Provisional Instrument for Assessing Clinical Practice Guidelines" (IOM instrument),[21] the "Method for Evaluating Research and Guideline Evidence" (MERGE instrument),[22] Cluzeau et al's "Appraisal Instrument for Clinical Guidelines" (Cluzeau instrument),[6] and Shaneyfelt et al's methodologic appraisal instrument (Shaneyfelt instrument).[3]

The first of these instruments, produced by the IOM Committee on Clinical Practice Guidelines, was designed to provide an explicit method of evaluating the soundness of clinical guidelines and a standardized method of assessing guidelines. The IOM instrument uses 142 items to evaluate guidelines for 8 attributes deemed by the IOM to be essential to usable practice guidelines.[21]

The Centre for Clinical Policy and Practice of the NSW Health Department (New South Wales, Australia) developed the MERGE instrument. It sets out an explicit, standardized approach to reviewing and incorporating scientific evidence into guidelines.[22]

The Cluzeau instrument is notable in that it was designed to be easy to use by physicians in the field. Moreover, its development did not include the elaborate review process associated with the IOM and MERGE tools. In addition, it provides a comparative scoring system. The instrument was developed in response to a call by the Committee on Clinical Practice Guidelines of the Institute of Medicine (part of the National Academy of Sciences) for the development of a generic guideline appraisal instrument. The Cluzeau instrument has been widely used throughout Europe and is currently being used by the National Health Service Executive in the United Kingdom to assist in guideline evaluation and recommendation.[6,15]

Table 1. *Cluzeau instrument: "Appraisal Instrument for Clinical Guidelines" (version 1)*

For each question, please answer "Yes," "No," "Not sure," or "Not applicable."

Dimension 1: Rigor of Development
- Responsibility for guideline development
  1. Is the agency responsible for the development of the guidelines clearly identified?
  2. Was external funding or other support received for developing the guidelines?
  3. If external funding or support was received, is there evidence that the potential biases of the funding body (or bodies) were taken into account?
- Guideline development group
  4. Is there a description of the individuals (eg, professionals, interest groups—including patients) who were involved in the guidelines development group?
  5. If so, did the group contain representatives of all key disciplines?
- Identification and interpretation of evidence
  6. Is there a description of the sources of information used to select the evidence on which the recommendations are based?
  7. If so, are the sources of information adequate?
  8. Is there a description of the method(s) used to interpret and assess the strength of the evidence?
  9. If so, is(are) the method(s) for rating the evidence satisfactory?
- Formulation of recommendations
  10. Is there a description of the methods used to formulate the recommendations?
  11. If so, are the methods satisfactory?
  12. Is there an indication of how the views of interested parties not on the panel were taken into account?
  13. Is there an explicit link between the major recommendations and the level of supporting evidence?
- Peer review
  14. Were the guidelines independently reviewed prior to their publication/release?
  15. If so, is explicit information given about methods and how comments were addressed?
  16. Were the guidelines piloted?
  17. If the guidelines were piloted, is explicit information given about the methods used and the results adopted?
- Updating
  18. Is there a mention of a date for reviewing or updating the guidelines?
  19. Is the body responsible for the reviewing and updating clearly identified?
- Overall assessment of development process
  20. Overall, have the potential biases of guideline development been adequately dealt with?

Dimension 2: Context and Content
- Objectives
  21. Are the reasons for the developing the guidelines clearly stated?
  22. Are the objectives of the guidelines clearly defined?
- Context
  23. Is there a satisfactory description of the patients to which the guidelines are meant to apply?
  24. Is there a description of the circumstances (clinical or nonclinical) in which exceptions might be made in using the guidelines?
  25. Is there an explicit statement of how the patient's preferences should be taken into account in applying the guidelines?
- Clarity
  26. Do the guidelines describe the condition to be detected, treated, or prevented in unambiguous terms?
  27. Are the different possible options for management of the condition clearly stated in the guidelines?
  28. Are the recommendations clearly presented?
- Likely costs and benefits
  29. Is there an adequate description of the health benefits that are likely to be gained from the recommended management?
  30. Is there an adequate description of the potential harms or risks that may occur as a result of the recommended management?
  31. Is there an estimate of the costs or expenditures likely to incur from the recommended management?
  32. Are the recommendations supported by the estimated benefits, harms, and costs of the intervention?

Dimension 3: Application
- Guideline dissemination and implementation
  33. Does the guideline document suggest possible methods for dissemination and implementation?
- Monitoring of guidelines/clinic audit
  34. Does the guideline document specify criteria for monitoring compliance?
  35. Does the guideline document identify clear standards or targets?
  36. Does the guideline document define measurable outcomes that can be monitored?
- National guidelines only
  37. Does the guideline document identify key elements that need to be considered by local guideline groups?

Instrument created by Françoise Cluzeau, Peter Littlejohns, Jeremy Grimshaw, and Gene Feder at St George's Hospital Medical School in May 1997. Reprinted with kind permission from the authors.

For a copy of the original document, contact: Health Care Evaluation Unit, Department of Public Health Sciences, St George's Hospital Medical School, Cranmer Terrace, London SW17 ORE. E-mail: *I.hall@sghms.ac.uk.*

The Shaneyfelt instrument was designed to be used by clinicians and researchers to assess guidelines. Although it has not undergone extensive testing, it has been evaluated and endorsed as valid by experts in the field of guideline construction. This tool uses 24 questions to evaluate guidelines.[2,3,32]

Procedures

The IOM instrument and the MERGE instrument provide for very thorough evaluations of guidelines and supporting evidence. Although we found both of them to be very helpful in providing in-depth insight regarding guideline quality, these instruments do not provide for comparative scoring systems; this makes comparative assessment of the guideline reviews difficult.

The Shaneyfelt and Cluzeau instruments were the only appraisal tools identified as having the ability to provide comparative scoring systems.[32] The Cluzeau instrument has been tested and described as a reliable and valid method of

Table 2. *Mercy and CCP dimension scores*

Guideline	Dimension 1	Dimension 2	Dimension 3	Total score
Mercy mean score	77.00	69.17	52.00	71.08
SD	7.89	14.72	21.50	9.19
CI (α = .05)	4.89	9.13	13.32	5.70
CCP mean score	20.00	10.83	0.00	14.32
SD	6.67	10.43	0.00	6.38
CI (α = .05)	4.13	6.46	—	3.95

guideline evaluation.[7,15,32] Graham et al[32] concluded that the Cluzeau instrument was the most well developed to date and that the Shaneyfelt instrument was the second most well developed. Only the Cluzeau instrument includes items that relate to each of the 10 attribute categories established by Graham et al.[32] Because it encompasses the recommendations made in the literature reviewed, has been recognized as valid by the scientific literature, provides a comparative scoring system, and is supported by an explanatory users' manual, we selected the Cluzeau instrument as the tool that we would use to evaluate the quality of the 2 chiropractic guidelines.

The Instrument

In response to the IOM's call for the development of a generic appraisal instrument, Cluzeau et al developed an appraisal instrument to assess the quality of clinical guidelines.[33] This appraisal tool was designed to assist in the identification of guidelines that lack rigorous development to minimize bias.[6] The instrument reviews each guideline in terms of 3 recognized attributes, or *dimensions*, of good guidelines in a checklist format.[6,12,22,23,30] Dimension 1 addresses the rigor of development, including evaluation of the development process and methods of identifying and interpreting evidence. Dimension 2 measures context and content, rating the guideline's readability, applicability, clarity, and focus on cost and benefit issues. Dimension 3 addresses dissemination, implementation, and application strategies for the guidelines. The resulting scores can be used to evaluate and compare guidelines. The data can also be used to assist guideline developers with revisions or with the development of updated guidelines. Part of the functionality of the Cluzeau instrument is that it can be used by physicians or groups on a grassroots level without the need of their going through the elaborate review process associated with the IOM and MERGE instruments. In addition, it is the only available instrument that has undergone testing for reliability and validity.[32]

The instrument has been found to have acceptable reliability, with a Cronbach α coefficient of 0.68 to 0.84 and an intraclass correlation coefficient of 0.82 to 0.92.[6] Because there is no gold standard with which to compare this instrument, its validity has been difficult to assess. Nevertheless, there are some data to support the instrument's validity. The instrument's authors report that the Pearson correlation coefficients between appraisers' dimensions and global assessment scores were 0.49 for dimension 1, 0.63 for dimension 2, and 0.40 for dimension 3. A research project known as BIOMED-2, which involved the utilization of this instrument in 10 European countries and Canada, further supported the tool's validity.[6,32] A European guideline appraisal tool called the "Appraisal of Guidelines, Research, and Evaluation in Europe" (AGREE) is currently under development and is based on the Cluzeau instrument.[32]

The Cluzeau instrument is shown in Table 1.[15]

Appraisal Instrument Application

The CCP and Mercy guideline documents were reviewed through use of methods of guideline evaluation derived from the search described above. Using the Cluzeau instrument, 10 volunteer appraisers performed independent evaluations of each of the guideline documents by assigning the document a score—"Yes," "No," "Not sure," or "Not applicable"—for each of the 37 criteria in the instrument. Each of the 10 appraisers had postgraduate training or experience in critical evaluation of scientific literature or research design. Version 1 of the "User's Guide" for the instrument was used to clarify criteria questions. For scoring purposes, 1 point was allotted for each criterion scored "Yes," and simple totals were calculated for each of the 3 dimensions and expressed as a percentage. These independent evaluations were tabulated, compared, and used to formulate a consensus on a final guideline evaluation synopsis. Mean scores, SDs, and confidence intervals were calculated. As a check of the internal validity, the reviewers separately scored each of the 2 sets of guidelines for its global assessment by reporting whether the guideline document was "strongly recommended for use in practice," "recommended for use with some modification or proviso," or "not recommended as suitable for use in practice." The relationship between each appraiser's global assessment and scoring was assessed with a Pearson correlation coefficient.

RESULTS

The total mean score for the CCP guidelines was 14.32%; this compared with a total mean score for the Mercy guidelines of 71.08%. Table 2 shows the distribution of scores for each dimension for the CCP and Mercy guidelines. Each of 7 criteria (35%) of dimension 1 are contingent on a "Yes" answer to another criterion in that dimension. Failure to qualify as "Yes" also results in failure to earn a point for that dependent criterion, resulting in significantly low scores for the CCP guidelines in both dimension 1 and total score, for which dependent criteria account for 19% of the possible points.

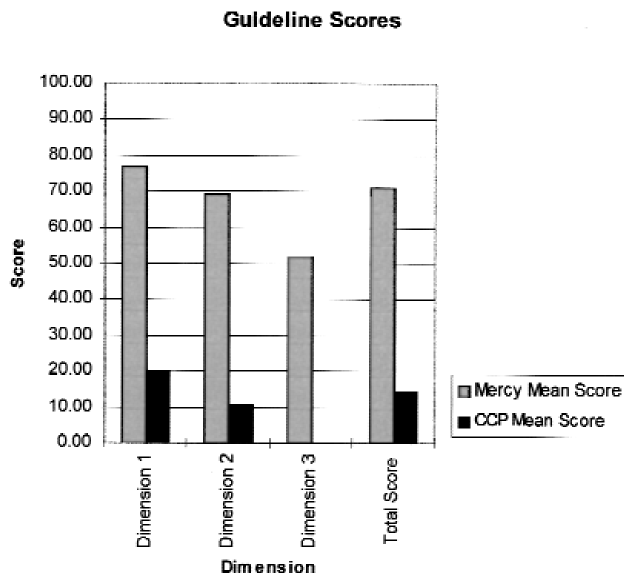


Fig 1. *Graphic comparison of Mercy and CCP guideline documents' quality assessment scores.*

Overall, appraisers agreed in responses to 85% of criteria applied to the Mercy guidelines and to 95% of criteria applied to the CCP guidelines. The disagreements pertaining to the CCP guidelines were for different criteria than those pertaining to the Mercy guidelines. Interappraiser agreement for individual dimensions, as measured by interclass correlation coefficient, was good (range, 0.84-0.939), and the mean dimension scores have small confidence intervals (4.13-9.13).

Appraisers rated the Mercy document as not specifying criteria for monitoring compliance, as not identifying clear standards or targets, and as not having been piloted. Although review and updating of the guidelines are mentioned in the Mercy document, the document did not satisfy the criterion asking for a date for that reassessment, nor does the Mercy document provide an estimate of the costs or expenditures likely to be associated with the recommended management.

Appraisers were unanimous that the CCP guidelines did not satisfy 81% of the possible criteria on the Cluzeau instrument. The CCP guidelines scored poorly in the area of guideline development because they do not include representation from several of the major interest groups within the chiropractic profession. Appraisers rated the CCP document as having failed to state the methods used to assess the quality and strength of the evidence selected for uses in the guidelines. In addition, the CCP guidelines failed to provide an acceptable rating scheme to support the quality of evidence for the given recommendations. The appraisers rated the CCP guidelines as having problems with context and clarity regarding application and recommendations. The CCP guidelines did not satisfy the criterion regarding a reassessment date, nor did they provide an estimate of the costs or expenditures likely to be associated with the recommended management.

Most (90%) of the appraisers scored the Mercy document as "recommended for use with some modification or proviso," and most (90%) scored the CCP document as "unacceptable."

The Pearson correlation coefficients between appraisers' dimensions scores and their global assessments were 0.68 for dimension 1 ($P < .001$), 0.54 for dimension 2 ($P = .014$), and 0.61 for total score ($P = .004$); the correlation between dimension 3 (measuring implementation, dissemination, and monitoring strategies) and global assessment was not significant ($r = 0.40$, $P = .077$).

The comparative scores of the guideline appraisals are shown in graph form in Fig 1. Narrative consensus guideline appraisal reports were generated for each guideline document.

DISCUSSION

The CCP guidelines are condition-specific and focus on the diagnosis and management of the vertebral subluxation. Within the chiropractic profession, there is debate as to the significance and existence of the vertebral subluxation. The CCP guidelines fail to objectively address both the scientific evidence and the controversy regarding the existence or significance of a subluxation as a diagnostic entity. It is not within the intent or scope of the present article to address this controversial issue, but the view that subluxation assessment is a valid diagnostic method is questioned in both the scientific literature[1,34,35] and those chiropractic guidelines indexed in the National Guideline Clearinghouse that have evidence rating systems suited to scientific and technical areas of practice.[11,36] Hayward et al[5] recommend that readers look for an explanation if new guidelines differ from those currently in use. It is also noted that those recommendations based solely on clinical judgment and experience are most susceptible to bias and self interest.[26] In the end, if the underlying evidence is weak, no matter what degree of consensus is attained, the validity of the guidelines will be limited.[5] Lack of methods to assess and report the quality and strength of scientific evidence considered, failure to consider relevant data, and over-reliance on clinical judgments and opinions appear to be the major flaws associated with the CCP guidelines. In addition, the CCP guidelines lack a rating system that is suited to scientific areas of practice; the document thereby prohibits the reader from assessing the evidence supporting the guideline recommendations. These flaws account for the poor appraisal scores noted for the CCP guidelines.

The Mercy guidelines were found to be both procedure-based and broad in scope and to have generally adhered to acceptable rigors of development. This was reflected in the final appraisal scores noted above. The major problem identified with the Mercy guidelines lies in the age of the document. Inasmuch as the Mercy document is based on the literature published before 1993, there are obviously several years of newer research to be considered. Nevertheless, the general consensus of the appraisers in this study was that the large majority of the recommendations found in the Mercy

guidelines continue to be supported by the current literature and that the guidelines are still valid and usable—with the proviso that current research findings published since the Mercy guideline's publication should be taken into consideration. According to the National Guideline Clearinghouse,[36] the Mercy guidelines were reaffirmed in 1999 by the Commission for the Establishment of Guidelines for Chiropractic Quality Assurance and Practice Parameters.

The greatest shortcoming of this study is that it is based on the utilization of an instrument that has not yet been established as a gold standard. Graham et al[32] note that currently there is insufficient evidence to support the exclusive use of any one appraisal instrument, although the Cluzeau instrument appears to be the one most well developed. Although the gold standard validity of the Cluzeau instrument remains to be established, it is considered the most valid and utilitarian tool of its kind available at this time. The appraisal tool was optimized to appraise condition-based guidelines rather than procedure-based guidelines; the Mercy scores may thus have been lower than they otherwise could have been, inasmuch as certain criteria did not apply well to the procedure-based Mercy document. The appraisal instrument does require a degree of subjective assessment, which is unavoidable in any evaluation tool.

Statistically, our results are within the value range of those of Cluzeau et al[6] and in their analysis of 60 guidelines. This fact supports the notion that this guideline appraisal instrument was applied and functioned as intended by its authors. Our independent appraisers reached a remarkable consensus supporting the reliability of this study's outcomes.

Ultimately, physicians and groups need to determine the quality of drafted guidelines, and the Cluzeau instrument facilitates such analysis. To avoid unwarranted utilization of poorly constructed guidelines, it is strongly recommended that the quality of all future guidelines be similarly reviewed for validity and scientific accuracy with the findings published in a medically indexed journal before adoption by any professional group.

CONCLUSION

Guidelines are considered valid when they optimize patient care and result in health gain at the expected costs. Procedures exist for guideline evaluation, and the utilization of such procedures can identify problems with guideline quality. At best, invalid guidelines can waste resources; at worst, they can be detrimental to patient health.[16] Early identification of guidelines lacking validity is essential so that professional acceptance of and reliance on flawed data can be avoided.

The Cluzeau instrument is a valid and useful method of guideline evaluation. It is recommended that future guidelines be evaluated for validity by means of an accepted appraisal tool, such as the one used in this review of the CCP and Mercy guideline documents.

A review of the results of this evaluation led us to conclude that the development process of the CCP document might be flawed partly as a result of bias with respect to the existence and significance of the subluxation and partly as a result of a flawed evaluation protocol and poor correlation of scientific evidence. According to the method of guideline evaluation used in this study, the CCP guidelines fail to document that fundamental steps of quality guideline development were followed. The notably low quality assessment scores for the CCP guidelines resulted in a general consensus that they are not suitable for use in chiropractic practice, and they received a rating of "Not recommended."

Reviewing the appraisal scores and evidence led us to conclude that the Mercy guidelines are valid and usable in chiropractic practice, with the proviso that current credible literature and research findings published since the Mercy document publication should be taken into consideration. On the basis of the results of our appraisal, the Mercy guidelines are "Recommended with proviso."

ACKNOWLEDGMENTS

We extend our thanks to Francoise Cluzeau for her helpful suggestions and assistance in reviewing this manuscript. We thank Phil Kashdan and the staff of the Oregon Public Library for their extraordinary help in obtaining needed journal articles and books. We also acknowledge Dale Hoppe for his valuable assistance in proofreading this manuscript.

REFERENCES

1. Meeker WC. The future impact of clinical practice guidelines. J Manipulative Physiol Ther 1995;18:606-10.
2. Cook D, Giacomini M. The trials and tribulations of clinical practice guidelines. JAMA 1999;281:1950-1.
3. 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.
4. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ 1999;318:527-30.
5. Hayward R, Wilson M, Tunis S, Bass EB, Guyatt G. Users' guides to the medical literature, VIII: how to use clinical practice guidelines, A: are the recommendations valid? The Evidence-Based Medicine Working Group. JAMA 1995;274:570-4.
6. Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Morgan S. Development and application of a generic methodology to assess the quality of clinical guidelines. Int J Qual Health Care 1999;11:21-8.
7. Feder G, Eccles M, Grol R, Griffiths C, Grimshaw J. Clinical guidelines: using clinical guidelines. BMJ 1999;318:728-30.
8. Fletcher SW, Fletcher RH. Development of clinical guidelines. Lancet 1998;352:1876.
9. Weingarten S, Riedinger M, Conner L, Johnson B, Ellrodt AG. Reducing lengths of stay in the coronary care unit with a practice guideline for patients with congestive heart failure: insights from a controlled clinical trial. Med Care 1994;32:1232-43.
10. Council on Chiropractic Practice. Vertebral subluxation in chiropractic practice. (Clinical practice guideline; no. 1.) Chandler (AZ): The Council; 1998.
11. Haldeman S, Chapman-Smith D, Petersen DM. Guidelines for chiropractic quality assurance and practice parameters: proceedings of the Mercy Center Consensus Conference. Gaithersburg (MD): Aspen Publishers; 1993.
12. Hayward RS, Wilson MC, Tunis SR, Bass EB, Rubin HR,

Haynes RB. More informative abstracts of articles describing clinical practice guidelines. Ann Intern Med 1993;118:731-7.
13. Brook RH. Implementing medical guidelines. Lancet 1995; 346:132.
14. Mootz RD, Shekelle PG. Practice guidelines, clinical pathways, and technology assessments in chiropractic. In: Cherkin D, Mootz R, editors. Chiropractic in the United States, practice, and research. Washington: Agency for Health Care Policy and Research; 1997. p. 72-96.
15. Cluzeau F, Littlejohns P, Grimshaw J, Feder G. Appraisal instrument for clinical guideline. London: St. George's Hospital Medical School; 1997.
16. Eccles M, Clapp Z, Grimshaw J, Adams PC, Higgins B, Purves I, et al. North of England evidence-based guidelines development project: methods of guideline development. BMJ 1996;312:760-2.
17. Fletcher JW, Woolf SH, Royal HD. Consensus development for producing diagnostic procedure guidelines: SPECT brain perfusion imaging with exametazime. J Nucl Med 1994;35: 2003-10.
18. Graham I, Beardall S, Carter A, Laupacis A. The state of the art of practice guidelines development, dissemination, and evaluation in Canada. Amsterdam, Netherlands: Scientific Basis for Health Services Conference; October 1997.
19. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet 1993;342:1317-22.
20. Grimshaw JM, Eccles M, Russell I. Developing clinically valid practice guidelines. J Eval Clin Pract 1995;1:37-48.
21. Institute of Medicine (US). Committee on Clinical Practice Guidelines. In: Field MJ, Lohr KN. Guidelines for clinical practice: from development to use. Washington: National Academy Press; 1992.
22. Liddle J, Williamson M, Irwig I. Method for evaluating research and guideline evidence. Sydney, Australia: NSW Health Department; 1996.
23. Lohr KN, Field MJ. A provisional instrument for assessing clinical practice guidelines. In: Institute of Medicine, editors. Guideline: from practice to use. Washington: National Academy Press; 1992.
24. Lohr KN. Reasonable expectations: from the Institute of Medicine (interview by Paul M. Schyve). QRB Quality Review Bulletin 1992;18:393-6.
25. Royal H, Pierson R Jr, Fletcher J, Dillehay G. Guidelines for guideline development. Society of Nuclear Medicine. J Nucl Med 1996;37:878-81.
26. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. BMJ 1999;318:593-6.
27. Steven I, Fraser R. Clinical practice guidelines: particular reference to the management of pain in the lumbosacral spine. Spine 1996;21:1593-6.
28. Woolf SH. Practice guidelines: a new reality in medicine, I: recent developments. Arch Intern Med 1990;150:1811-8.
29. Woolf SH. The process of developing practice guidelines. Decubitus 1991;4:28-31.
30. Woolf SH. Practice guidelines, a new reality in medicine, II: methods of developing guidelines. Arch Intern Med 1992;152: 946-52.
31. Woolf SH, DiGuiseppi CG, Atkins D, Kamerow DB. Developing evidence-based clinical practice guidelines: lessons learned by the US Preventive Services Task Force. Annu Rev Public Health 1996;17:511-38.
32. Graham ID, Calder LA, Hébert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. Int J Technol Assess Health Care 2000;16:1024-38.
33. Field MJ, Lohr KN, Institute of Medicine (US). Committee to Advise the Public Health Service on Clinical Practice Guidelines. United States Department of Health and Human Services. Clinical practice guidelines: directions for a new program. Washington DC: National Academy Press; 1990.
34. Keating JC Jr. To hunt the subluxation: clinical research considerations. J Manipulative Physiol Ther 1996;19:613-9.
35. Leboeuf-Yde C. How real is the subluxation? a research perspective. J Manipulative Physiol Ther 1998;21:613-9.
36. National Guideline Clearinghouse. Guideline comparison. Sponsored by the Agency for Health Care Policy and Research in partnership with the American Medical Association and the American Association of Health Plans. Washington DC: The Agency; 1999.