

Evaluation of clinical practice guidelines in chiropractic care: a comparison of North American guideline reports

Melissa Brouwers, PhD*†
 Manya Charette, BSc*†

Clinical practice guidelines developed by the Canadian Chiropractic Association (CCA) and the Council on Chiropractic Practice (CCP) were evaluated by three independent appraisers using the most current version of the Appraisal of Guidelines for Research and Evaluation in Europe (AGREE) Instrument. Eighteen eligible chapters within the two documents (nine from each organization) were evaluated. In general, the CCA document was rated more favourably than the CCP document. The strengths of both documents include clarity of objectives and target users and complete descriptions of methods used to formulate recommendations. Areas of improvement for both documents include the need for more detail regarding the bodies of evidence under consideration for each section of the guideline. This paper presents the complete results of the evaluation, discusses the strengths of each guideline document, and makes suggestions for areas of improvement.

(JCCA 2001; 45(3):141-153)

KEY WORDS: guidelines, practice.

Les guides de pratique clinique élaborés par l'Association chiropratique canadienne (ACC) et par le Council on Chiropractic Practice (CCP) ont été évalués par trois experts indépendants au moyen de la version la plus récente de l'instrument Appraisal of Guidelines for Research and Evaluation in Europe (AGREE). Dix-huit chapitres admissibles dans les deux documents (neuf de chaque organisme) ont fait l'objet d'une évaluation. En général, le document de l'ACC a été mieux coté que celui du CCP. Parmi les forces des deux documents, on note la clarté de l'énoncé des objectifs et des utilisateurs cibles ainsi que le caractère exhaustif des descriptions de méthodes employées pour la formulation de recommandations. Les points à améliorer, dans les deux cas, comprennent le besoin de détailler l'accumulation des preuves à l'étude pour chaque section des guides. Cet article présente les résultats complets de l'évaluation, examine les points forts de chaque guide et offre des suggestions concernant les points à améliorer.

(JACC 2001; 45(3):141-153)

MOTS CLÉS : guides, pratique.

* Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

† Program in Evidence-based Care, Cancer Care Ontario, Hamilton, Ontario, Canada.

Corresponding author: Melissa Brouwers, PhD, McMaster University Health Sciences Centre, Room 3H7A, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5. Tel.: 905-525-9140, ext. 22055. mbrouwer@mcmaster.ca

Project sponsored by the Canadian Chiropractic Association.

© JCCA 2001.

INTRODUCTION

Clinical practice guidelines are systematically developed statements designed to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.¹ In the area of chiropractic practice, two key clinical practice guideline documents have been produced in North America: Clinical Practice Guidelines for Chiropractic Practice in Canada commissioned by the Canadian Chiropractic Association (CCA),² and the American counterpart, Vertebral Subluxation in Chiropractic Practice, Clinical Practice Guideline by the Council on Chiropractic Practice (CCP).³ The aim of both guideline documents is to facilitate an evidence-based approach to chiropractic care that enables conscientious, explicit and judicious use of the current best evidence in making decisions about the care of the individual patient.⁴

In 2000, a proposal was submitted and accepted by the CCA to evaluate the CCA and CCP clinical practice guidelines. The relative strengths and weaknesses between the guideline documents and within each document were considered, as well as their adherence to evidence-based principles.

The documents were evaluated using the most current version of the Appraisal of Guidelines for Research and Evaluation in Europe (AGREE) Instrument.⁵ Although there is no universally accepted, validated tool for measuring the quality of practice guidelines, the AGREE instrument has been widely studied and is currently undergoing international validation as part of the *BIOMED-PL96-3669* project. One of the appraisers (MB) is a member of this collaboration and special permission was granted by Francoise Cluzeau, project coordinator of *BIOMED-PL96-3669*, to use the most current version of the instrument (May 1999). The AGREE instrument is generic and can be applied to practice guidelines that cover topics such as prevention, diagnosis, treatment or intervention in any disease area.

The appraisers are methodology specialists who have no experience or training with chiropractic practice or the specific clinical and policy literature in this area. The scope of this evaluation focused exclusively on the chapters in each of the guideline documents that met the criteria for appraisal set out by the instrument, with the emphasis directed towards the methodological quality of the documents. The evaluation did not include the appraisal of clinical care and policy considerations.

METHODS

Appraisers

The research team mandated to review the clinical practice guideline documents developed by the CCA and CCP was composed of three independent appraisers. The appraisers have expertise in the development of clinical practice guidelines for cancer care and in the implementation of the evaluation tool used in this study. The appraisers are methodological experts with training and experience in systematic review, critical appraisal, and study design. One of the appraisers (MB), a doctorate-level behavioural scientist, coordinated the overall project.

This initiative was funded by the CCA. None of the appraisers have had any previous professional relationships with the CCA or the CCP. The CCA and the project coordinator agreed upon the scope of the evaluation before the initiative began. The research team had complete editorial independence from the CCA in the execution of the evaluation process, the final report submitted to the CCA, and the writing of this paper.

Evaluation instrument

It was agreed by the CCA and the project coordinator that the documents would be evaluated using the most current version of the AGREE Instrument.⁵ The AGREE Instrument capitalizes on the large literature identifying attributes that define high quality clinical practice guidelines (CPGs).^{6,7,8,9,10} Its purpose is to provide a framework to assess the confidence that the potential biases (e.g., methodological biases^{11,12,13}) in guideline development have been adequately addressed, that the recommendations that emerge from the CPG are reliable and valid, and that practical issues have been addressed.

It is composed of 24 items organized into eight dimensions. The *scope and purpose* dimension consists of four items and considers definitions and descriptions of guideline objectives, target users, clinical questions and patient population. The *stakeholder involvement* dimension is measured by four items that address the membership of the guideline development group, external review of the guidelines, pilot testing, and the extent of patient involvement in the development of the guidelines. The dimension, *identification and use of the evidence*, consists of two items and considers the methods used to search for the evidence base and the criteria for selecting the evidence.

The *formulating recommendations* dimension is measured by four items and is concerned with the formulation of recommendations, links between the evidence and recommendations, consideration of possible benefits and risks, and the impact on resources. The *clarity* dimension consists of four items and addresses the guideline structure, recommendations, and options for care. The *application* dimension is focused on organizational barriers, attitude/behaviour change issues and tools for application, and consists of three items. *Monitoring*, consisting of two items, focuses on the criteria for adherence to the recommendations and the updating process. The *editorial independence* dimension consists of one item. The items are answered using a combined ordinal-dichotomous scale composed of six response options that includes a 4-point likert scale (points *strongly agree*, *agree*, *disagree*, *strongly disagree*) and two additional options *no information to answer* and *not applicable*.^a

Scope of the evaluation

The project coordinator reviewed each of the guideline documents prior to beginning the evaluation. Each chapter was categorized as either eligible for evaluation according to the AGREE criteria, ineligible for evaluation according to the AGREE criteria, or as supporting (see Table 1). Eligible chapters were those that focused on specific clinical care issues (e.g., diagnosis, treatment). Where possible, each of the eligible chapters from one document was paired with an equivalent chapter from the second document (e.g., *CCA Frequency and Duration of Care* was paired with *CCP Duration of Care for Vertebral Subluxation*). The AGREE instrument was completed for each eligible chapter. Chapters that focused on non-clinical issues or issues outside the scope of the AGREE instrument were categorized as ineligible, and were not considered in the evaluation. Methodological issues such as the quality of the systematic review and adherence to evidence-based principles are essential components to high quality CPGs and are reflected in the AGREE instrument.^{6,7,9,14} These issues were often represented in what we have termed supporting chapters in the CCA and CCP guideline documents. Thus, for some items in the AGREE instrument, a general score was derived using the supporting chapters as a foundation on which the appraisers' response options were based. In the absence of additional information within the eligible chapters in each guideline document,

the general scores for these items were applied. The evaluation focused on the written documents supplied by the CCA and CCP. No efforts were made by the evaluators to verify the processes, participation and methods outlined in the guideline documents.

Procedures

The research team initially met to review each of the items in the AGREE instrument, to clarify any ambiguities with item content, and to further operationally define items where, at face value, potential misunderstandings or inconsistencies in interpretation were possible. At the conclusion of that meeting, each appraiser was provided with copies of the CCA and CCP guideline documents, a list outlining the categories in which the chapters were placed (Table 1), and 18 copies of the AGREE instrument. The AGREE instrument was completed for the nine eligible chapters from each organization. For both the CCA and CCP documents, the appraisers first read the supporting chapters, followed by the first eligible chapter, and then completed the AGREE assessment tool. A consultation meeting followed this process to review the results of the first pair of assessments, to address inconsistencies in the application of questionnaire items, to further refine the operational definitions of problematic items, and to determine which of the instrument items could be classified as general.

Appraisers then completed a review and assessment of each eligible chapter pair as outlined in Table 1. The order of evaluation within each pair was counterbalanced to avoid bias based on chapter order. Consultation meetings were held after the first, second and third chapter pairs.^b

Treatment of the data

Scoring

To evaluate the strengths and weaknesses of the CCA and CCP documents, responses provided by each of the assessors for each of the guidelines were obtained. *Strongly agree*, *agree*, *disagree* and *strongly disagree* responses were scored 4, 3, 2, and 1, respectively. *No information to*

^a For four items, only the 4-point likert scale was available.

^b One appraiser was not available for the second consultation meeting; the remaining appraisers reviewed her results.

Table 1
Categorization of Canadian Chiropractic Association (CCA)
and Council on Chiropractic Practice (CCP) guidelines and order (O) of evaluation

	CCA	CCP	Category
1	Chapter 3: Initial Clinical Examination	Chapter 1. History and Chiropractic Exam	Eligible
2	Chapter 4. Diagnostic Imaging	Chapter 3. Radiographic and Support Instrumentation	Eligible
3	Chapter 6. Diagnostic Instrumentation	Chapter 2. Instrumentation	Eligible
4	Chapter 7. Clinical Impression and Diagnosis	Chapter 4. Clinical Impression and Assessment	Eligible
5	Chapter 8. Frequency and Duration of Care	Chapter 7. Duration of Care for Vertebral Subluxation	Eligible
6	Chapter 9. Reassessment	Chapter 5. Reassessment and Outcome Assessment	Eligible
7	Chapter 10. Modes of Care and Management	Chapter 6. Modes of Adjustive Care.	Eligible
8	Chapter 11. Outcome Assessment	Chapter 5. Reassessment and Outcome Assessment (review complete)	Eligible
9	Chapter 13. Contraindications and Complications	Chapter 9. Patient Safety	Eligible
10	No chapter equivalent.	Chapter 8. Chiropractic Care of Children	Eligible
REVIEWED WITH ORIGINAL GUIDELINE PAIR	Review Group	Contributors and Panel Members	Supporting
	Introduction and Guide to Use of These Guidelines	Introduction and Methodology	Supporting
	Chairman's Preface	No chapter equivalent	Supporting
	Commission's Letter	No chapter equivalent	Supporting
	President's Letter	No chapter equivalent	Supporting
	Funding and Sponsoring Organizations	No chapter equivalent	Supporting
	Steering Committee	No chapter equivalent	Supporting
	Commission	No chapter equivalent	Supporting
	Staff	No chapter equivalent	Supporting
	Consultants	No chapter equivalent	Supporting
	Chapter 16. Guidelines for the Development and Implementation of Practice Guidelines	No chapter equivalent	Supporting
	No chapter Equivalent	Peer Reviewers	Supporting
/	Chapter 1: Informed Consent	No chapter equivalent	Ineligible
/	Chapter 2: Record Keeping	No chapter equivalent	Ineligible
/	Chapter 5. Clinical Laboratory Procedures	No chapter equivalent	Ineligible
/	Chapter 12. Professional Relations	No chapter equivalent	Ineligible
/	Chapter 14. Preventative/Maintenance Care	No chapter equivalent	Ineligible
/	Chapter 15. Continuing Education and Professional Development	Chapter 10. Professional Development	Ineligible
/	Chapter 17. Practice Advertising.	No chapter equivalent	Ineligible

answer responses were given a score of 0, with the assumption that each of the characteristics featured in the instrument contributes to the overall quality of the document. Thus, if there were no indication that the characteristic was featured in the process, this would reflect poorer quality. Further, a *not applicable* response was also given a score of 0. However there was only one occasion in which the appraisers used this latter option.

Analyses

Data were analyzed using SPSS Version 10.0.5 for Windows.

Interrater reliability

The Kappa coefficient is a reasonable method for calculating the interrater reliability of categorical data, whereas the intraclass correlation coefficient (ICC) is an appropriate measure when continuous data are used.¹⁵ One of the challenges of a combined ordinal/categorical scale is determining the most appropriate method for calculating interrater reliability. However, given the scoring rationale described above, the fact that four of the items included only the ordinal scale component, and that the *not applicable* option was chosen only once by the assessors, it was decided that an ICC would be the most appropriate and best estimate measure of interrater reliability.

Evaluation

Various evaluation scores were calculated. First, item scores across the appraisers were summed to give a total score for each dimension and the entire questionnaire. Second, these scores were compared to the maximum total score (mts). The mts was calculated by multiplying the number of appraisers by the number of items in the instrument component under consideration (e.g., dimension or complete questionnaire) by the highest possible score (i.e., *strongly agree* response with a score of 4). For example, the mts for the *scope and purpose* dimension is 48 (3 appraisers \times 4 items in *scope and purpose* dimension \times 4 highest possible score). Third, score means and standard deviations were calculated for each of the eight dimensions and the entire questionnaire for the eligible chapters across each of the appraisers. These values were calculated for the whole CCA and CCP guideline documents as well as each eligible chapter within these two documents.

Finally, rank order of the total scores of the guideline chapters is presented.

RESULTS

Interrater reliability and differences between raters' scores

Measures of interrater reliability across all items and chapters revealed an ICC of $r = 0.76$ (95% confidence intervals [CI] = 0.73–0.79) for the CCA guidelines and $r = 0.77$ (95% CI = 0.74–0.80) for the CCP guidelines. Thus, adequate reliability was achieved.¹⁵

A one-way analysis of variance was conducted to determine if there were statistically significant differences among the appraisers on mean evaluation scores across all items and all chapters in each practice guideline document. For the CCA document, a statistically significant main effect for appraisers was found, $F(2,645) = 8.23$, $p < 0.001$. The mean evaluation score was more positive for one of the appraisers (reviewer B: $m = 2.78$, $sd = 1.09$), than either of the others (Reviewer A: $m = 2.31$, $sd = 1.34$ and Reviewer C: $m = 2.52$, $sd = 1.24$). Although the means fell in a similar pattern, no statistically significant difference among appraisers on mean evaluation scores for the CCP guidelines was found, $F(2,645)$, $p = 0.12$ (Reviewer A: $m = 1.76$, $sd = 1.39$; Reviewer B: $m = 2.00$, $sd = 1.45$; Reviewer C: $m = 1.84$, $sd = 1.49$). To increase reliability and address the biases in scoring, score means were used for the inferential statistics.

Evaluation scores

Tables 2 and 3 provide quantitative summaries of the evaluation results. Table 2 summarizes the distribution of responses for the CCA and CCP documents. Table 3 includes the mean, standard deviation, and total scores as a function of instrument component (i.e., eight dimensions and the complete instrument) for each of the eligible chapters and the guideline documents as a whole. Also included in Table 3 are the maximum total score (mts) as a function of instrument component.

Distribution of scores

A chi-square analysis was conducted to determine if the distribution of responses reported in Table 2 was significantly different between the two groups (CCA v. CCP). A significant difference was found, $X^2(4) = 138.44$, $p < 0.001$. As can be seen in Table 2, both groups had a similar proportion of *strongly agree* responses (CCA = 18.5% v. CCP = 16.8%). In contrast, the evaluation re-

sulted in two and a half times as many *agree* responses for the CCA document relative to the CCP document (47.5% v. 19.4%), over 40% fewer *disagree* responses (14.4% v. 22.5%) and half as many *strongly disagree* and *no information to answer* responses (CCA 6.6% and 13.0% v. CCP 13.3% and 27.9%, respectively). Interestingly, the most common response option in the CCA evaluation was *agree* (47.5%) in contrast to the *no information to answer* response with the CCP evaluation (27.5%). For both groups, the *strongly disagree* response was the least common response option.

Chapter score evaluation

Based on these findings, it is not surprising that the mean scores and total scores suggest a tendency for the eligible chapters in the CCA document to be rated more highly than the corresponding chapters in the CCP document. Although the ranges were the same, the minimum to maximum of total scores for eligible chapters in the CCA document were higher (171 to 201) and did not overlap with the

CCP document scores (115 to 145). Table 4 summarizes the rankings of all eligible chapters based on the total scores of the instrument (including those with no chapter equivalent). There are two noteworthy features of these rankings. First, the total scores of the CCA chapters were consistently higher (Tables 3, 4). Second, there appears to be no correspondence in the rankings of the CCA and CCP chapter pairs. For example, the strongest eligible CCA chapter was *Frequency and Duration of Care*. Its CCP counterpart, *Duration of Care for Vertebral Subluxation* was one of the organization’s weakest chapters (Table 4). In fact, the biggest difference in total scores and mean scores was found with this chapter pair. Similarly, the CCP counterpart of one of the weakest CCA chapters, *Modes of Care and Management*, was one of its strongest chapters, *Modes of Adjustive Care*. Although the total score of the CCA version of this chapter was still higher than the CCP version, the relative ranked positions of these chapters were very different.

Table 2
Distribution of response options of the Canadian Chiropractic Association (CCA)
and Council for Chiropractic Practice (CCP) guideline documents

Response Option	CCA		CCP	
	frequency†	percentage	frequency†	percentage **
<i>no information to answer/NA*</i> scored 0	84	13.0	181	27.9
<i>strongly disagree</i> scored 1	43	6.6	86	13.3
<i>disagree</i> scored 2	93	14.4	146	22.5
<i>agree</i> scored 3	308	47.5	126	19.4
<i>strongly agree</i> scored 4	120	18.5	109	16.8
	648	100.0	648	99.9

* NA: Not Applicable. This option was used by 3 appraisers for the CCA Contraindications and Complications chapter for the item addressing health benefits, risks and side effects.

** Please note that this column does not add to 100 % because of rounding.

† The frequency is out of a total of 648 possible scores (24 items on the instrument × 3 appraisers × 9 chapters).

Table 3
Dimension and complete questionnaire mean, standard deviation and total scores
for each applicable chapter from the Canadian Chiropractic Association (CCA)
and Council for Chiropractic Practice (CCP) practice guideline reports

ORGANIZATION AND CHAPTER TITLE	DATA	SCOPE & PURPOSE	STAKEHOLDER INVOLVEMENT	ID AND USE OF EVIDENCE	FORMULATING RECOMMENDS.	CLARITY	APPLICATION	MONITORING	EDITORIAL INDEPENDENCE	COMPLETE QUESTIONNAIRE
		<i>mts = 48</i>	<i>mts = 48</i>	<i>mts = 24</i>	<i>mts = 48</i>	<i>mts = 48</i>	<i>mts = 36</i>	<i>mts = 24</i>	<i>mts = 12</i>	<i>mts = 288</i>
CCA Initial Clinical Exam	m	2.7	2.1	1.7	2.4	3.0	2.7	2.5	0	171
	sd	0.49	1.38	0.82	1.08	0.95	1.58	1.38	0	
	total	32	25	10	29	36	24	15	0	
CCP History and Chiro- practic Exam	m	2.3	2.8	1.0	2.0	3.1	0.3	1.2	0	137
	sd	0.97	1.71	0.00	1.28	1.51	1.00	1.33	0	
	total	27	33	6	24	37	3	7	0	
CCA Diagnostic Imaging	m	3.3	2.1	2.0	3.2	2.8	3.4	3.0	0	197
	sd	0.45	1.38	0.63	0.39	0.58	0.53	0.63	0	
	total	39	25	12	38	34	31	18	0	
CCP Radiographic and Other Instrumentation	m	2.6	2.8	1.0	2.4	3.0	0.3	1.2	0	145
	sd	0.90	1.71	0.00	0.51	0.74	1.00	1.33	0	
	total	31	33	6	29	36	3	7	0	
CCA Diagnostic Instrumentation	m	3.2	2.1	1.7	2.6	2.8	2.7	2.5	0	177
	sd	0.58	1.38	0.82	1.31	0.39	1.58	1.38	0	
	total	38	25	10	31	34	24	15	0	
CCP Instrumentation	m	3.3	2.3	1.0	1.5	3.2	0.3	1.2	0	139
	sd	0.75	1.83	0.00	1.17	0.72	1.00	1.33	0	
	total	39	28	6	18	38	3	7	0	
CCA Clinical Impression and Diagnosis	m	2.6	2.1	1.7	2.2	3.4	2.7	2.5	0	172
	sd	0.67	1.38	0.82	1.19	0.79	1.58	1.38	0	
	total	31	25	7	26	41	24	10	0	
CCP Clinical Impression and Assessment	m	2.4	2.5	1.0	1.9	2.9	0.6	1.2	0	135
	sd	0.90	1.73	0.00	1.24	0.29	1.13	1.33	0	
	total	29	30	6	23	35	5	7	0	
CCA Frequency and Duration of Care	m	3.8	2.1	1.7	3.0	3.4	3.2	2.5	0	201
	sd	0.45	1.38	0.42	0.74	0.67	0.67	1.38	0	
	total	45	25	10	36	41	29	15	0	
CCP Duration of Care for Vertebral Subluxation	m	2.8	2.4	1.0	1.4	2.1	0.3	1.2	0	121
	sd	0.94	1.83	0.00	1.00	1.00	1.00	1.33	0	
	total	34	29	6	17	25	3	7	0	
mts: maximum total score		M: grand mean								
m: mean		SD: standard deviation of grand mean								
sd: standard deviation										

Table 3 Continued

ORGANIZATION AND CHAPTER TITLE	DATA	SCOPE & PURPOSE <i>mts = 48</i>	STAKEHOLDER INVOLVEMENT <i>mts = 48</i>	ID AND USE OF EVIDENCE <i>mts = 24</i>	FORMULATING RECOMMENDS. <i>mts = 48</i>	CLARITY <i>mts = 48</i>	APPLICATION <i>mts = 36</i>	MONITORING <i>mts = 24</i>	EDITORIAL INDEPENDENCE <i>mts = 12</i>	COMPLETE QUESTIONNAIRE <i>mts = 288</i>
CCA Reassessment	m	3.3	2.1	1.7	2.1	3.1	2.9	2.5	0	178
	sd	0.49	1.38	0.82	1.31	1.24	1.27	1.38	0	
	total	40	25	10	25	37	26	15	0	
CCA Outcome Assessment	m	3.2	2.1	2.2	3.0	3.2	2.7	2.5	0	182
	sd	0.58	1.38	0.98	0.60	0.58	1.58	1.38	0	
	total	38	25	13	36	38	24	15	0	
CCP Reassessment and Outcome Assessment	m	2.5	2.5	1.0	1.5	2.5	0.3	1.0	0	123
	sd	0.90	1.73	0.00	0.90	0.67	1.00	1.10	0	
	total	30	30	6	18	30	3	6	0	
CCA Modes of Care & Management	m	3.0	2.1	1.7	2.7	2.5	2.6	2.5	0	171
	sd	0.60	1.38	0.82	0.78	0.52	1.51	1.38	0	
	total	36	25	10	32	30	23	15	0	
CCP Modes of Adjustive Care	m	3.3	2.8	1.0	1.5	2.7	0.7	1.2	0	141
	sd	0.75	1.71	0.00	1.00	1.07	1.32	1.33	0	
	total	39	33	6	18	32	6	7	0	
CCA Contraindications and Complications	m	3.4	2.1	1.8	2.6	2.7	2.7	2.5	0	172
	sd	0.51	1.38	0.75	1.16	1.72	1.58	1.38	0	
	total	41	25	9	31	32	24	10	0	
CCP Patient Safety	m	2.4	2.3	1.0	1.3	2.4	0.3	1.2	0	115
	sd	1.08	2.01	0.00	0.98	1.08	1.00	1.33	0	
	total	29	27	6	16	29	1	7	0	
CCP Chiropractic Care for Children	m	2.8	2.6	1.0	1.6	2.4	0.6	1.2	0	130
	sd	1.22	1.73	0.00	1.00	1.08	1.13	1.33	0	
	total	33	31	6	19	29	5	7	0	
		<i>mts = 432</i>	<i>mts = 432</i>	<i>mts = 216</i>	<i>mts = 432</i>	<i>mts = 432</i>	<i>mts = 324</i>	<i>mts = 432</i>	<i>mts = 108</i>	<i>mts = 2592</i>
CCA Practice Guidelines Grand CCA mean = 2.3 Grand CCA SD = 0.02	m	3.2	2.1	1.8	2.6	3.0	2.8	2.6	0	1461
	sd	0.38	1.38	0.63	0.51	0.40	0.57	0.81	0	
	total	309	200	84	258	282	205	123	0	
CCP Practice Guidelines Grand CCP mean = 1.5 Grand CCP SD = 0.02	m	2.7	2.5	1.0	1.7	2.7	0.4	1.2	0	1042
	sd	0.45	0.34	0.00	0.41	0.43	0.57	0.23	0	
	total	260	241	48	153	255	29	55	0	
Overall Means and Standard Deviations	M	2.9	2.3	1.4	2.2	2.8	1.6	1.9	0	
	SD	0.42	0.27	0.42	0.60	0.37	1.26	0.73	0	
mts: maximum total score m: mean sd: standard deviation					M: grand mean SD: standard deviation of grand mean					

Instrument dimension evaluation

General findings

For the overall CCA document, mean scores on the dimensions from strongest to weakest are *scope and purpose* (3.2), *clarity* (3.0), *application* (2.8), *formulating recommendations* (2.6), *monitoring* (2.6), *stakeholder involvement* (2.1), *identification and use of evidence* (1.8) and *editorial independence* (0). For the overall CCP document, the mean scores are *scope and purpose* (2.7) and *clarity* (2.7), *stakeholder involvement* (2.5), *formulating recommendations* (1.7), *monitoring* (1.2), *identification and use of evidence* (1.0), *application* (0.4), and *editorial independence* (0). It is interesting to note that the *scope and purpose* and *clarity* dimensions were the strongest dimensions for both the CCA and CCP. Similarly, *identification and use of evidence* and *editorial independence* were two of the weakest areas for both groups.

DISCUSSION

The purpose of this project was to evaluate the CCA and CCP clinical practice guideline documents using the AGREE instrument.⁵ The findings from the evaluation process indicate that the CCA guideline document was rated more favourably than the CCP guideline document.

For both the CCA and CCP document, the least frequent response option was *strongly disagree*. However, for the CCA document the *agree* option was most frequent in contrast to the *no information to answer* option for the CCP document. Indeed, recall that we were unable to find information to answer over one quarter of the quality items related to the CCP document. The interval of total scores for eligible chapters in the CCA document was higher (171 to 201) and did not overlap with the interval of total scores in the CCP document (115 to 145). Table 4 presents the rank ordering of the eligible chapters in the CCA and CCP guideline documents. There appears to be little corre-

Table 4
The rank ordering of eligible chapters from the Canadian Chiropractic Association (CCA) with the corresponding eligible chapters from the Council for Chiropractic Practice (CCP) practice guideline reports according to total score

CCA			CCP		
Total Score	Rank	Chapter	Total Score	Rank	Corresponding Chapter
201	1	Frequency and Duration of Care	121	8	Duration of Care for Vertebral Subluxation
197	2	Diagnostic Imaging	145	1	Radiographic and Other Instrumentation
182	3	Outcome Assessment	123	7	Reassessment and Outcome Assessment
178	4	Reassessment			
177	5	Diagnostic Instrumentation	139	3	Instrumentation
172	6.5	Contraindications and Complications	115	9	Patient Safety
172	6.5	Clinical Impression and Diagnosis	135	5	Clinical Impression and Assessment
171	8.5	Modes of Care and Management	141	2	Modes of Adjustive Care
171	8.5	Initial Clinical Examination	137	4	History and Chiropractic Exam
		no chapter equivalent	130	6	Chiropractic Care of Children

spondence in the rank order of common-theme chapters between the two groups, suggesting that the subject matter of the guideline chapters did not consistently predict quality scores.

The CCA chapter scores were typically higher on the *scope and purpose* dimension than the CCP chapters, with the exception being the chapters that addressed modality of care and instrumentation (Table 3). Typically, the differences between the organizations were modest, with the exception of the chapters that addressed safety and complications (i.e., CCA *Contradictions and Complications* and CCP *Patient Safety*), where there was a full point difference between the mean scores. The differences between the two groups were attributed primarily to the inclusion of additional information regarding clinical questions of interest and patient characteristics in the CCA chapter.

The CCP document scores were consistently higher than the CCA scores on the *stakeholder involvement* dimension due primarily to the greater involvement of patients and methods experts in the guideline development process (Table 3). The variation within the CCP document can be attributed primarily to the extent to which patient preferences were considered in each chapter. The CCA score was the same across all chapters because information used to answer the items in this section was found in the supporting literature.

The primary area for improvement for both the CCA and CCP guidelines falls within the *identification and use of evidence* dimension. Both the CCA and the CCP scored quite low on this dimension. Neither group provided comprehensive descriptions of the methods used to search, select, and synthesize the evidence (Table 3).

The mean scores for *formulating recommendations* were higher for the CCA document compared to the CCP document, with the former providing more explicit and complete descriptions. At no time was the mean score of the CCP chapter higher than its CCA counterpart on this dimension.

The *clarity* dimension is one of the highest ranked dimensions for both the CCA and CCP documents. With the exception of chapters dealing with the duration of care, where the difference in means is greater than one point (CCA = 3.4 v. CCP = 2.1), the scores are very close between the groups (Table 3).

The *application* dimension yields the greatest and most

dramatic difference in scores between the CCA and CCP documents (Table 3). This can be attributed to the CCA document addressing some theoretical aspects of application and the inclusion of flow charts and algorithms for the user. There was no overlap between the groups on mean score ranges.

As with the *application* subdimension, substantial differences on scores between the two groups are found with the *monitoring* dimension (Table 3). The CCA document out performed the CCP document and there was no overlap between the mean ranges.

Finally, there was considerable debate among the appraisers regarding *editorial independence*. Financial support to develop the guidelines for the CCA document comes from professional associations and groups that would have an interest in the results of the project. Members of these organizations were involved in each step of the development process. The financial supporters of the CCP activities are unknown. In the absence of an explicit statement indicating editorial independence from the funding body, it was felt there was insufficient information to answer this question positively for either development group. The means and total scores for both the CCA and CCP document are zero for all of the chapters.

An important consideration is whether the differences in quality favouring the CCA document are meaningful and important. Three points bear on this issue. First, the greatest differences in total scores between the CCA and CCP documents are with the dimensions *application* (205 v. 29), *formulating recommendations* (258 v. 153), and *monitoring* (123 v. 55). Within these dimensions, the differences can be attributed, in large part, to the absence of information in the CCP guideline document rather than a description indicating that the guideline developers undertook a faulty or weak methodology. The distinction between poor guideline process and poor reporting standards is important. Based on the argument above, there is at least evidence to support the notion that the CCA reporting style is more complete than the CCP reporting style.

It is anticipated that this style of presentation, as it was used in the CCP document, was purposeful. In the introductory statement, the CCP document indicated a commitment to making a “user friendly” compendium. Indeed, the actual text of the document was short and the use of bold typeface successfully highlighted the recommendations. However, this streamlining approach by the CCP may

have inadvertently neglected the inclusion of very important information that would enable the reader to make informed assessments about the recommendations. The “user friendly” compendium perhaps could have served as a supporting document to one in which all the issues were addressed.

The second aspect that bears on this issue is related to perceptions about the relevance of the individual dimensions. There may be some debate among the guideline development groups regarding the relevance of each dimension of the AGREE instrument within the clinical, policy, and health services contexts in the chiropractic community. For example, the greatest discrepancies between the CCA and CCP documents were with the *application* dimension and the *monitoring* dimension. If these dimensions are considered less relevant by the chiropractic community than other dimensions in which there is greater correspondence, it could be concluded that the differences in scores between the two reports might be less meaningful than believed at first glance. Many factors can influence this debate: the guideline model of the development groups, the mandate and responsibilities of development groups in contrast to the mandate and responsibilities of the professional organizations, the expertise of the members of the groups (e.g., clinical experts vs. methodology experts vs. implementation experts, etc.).

Finally, examining response patterns is the third consideration when trying to understand the evaluation outcome differences between the two groups. The CCA document received twice as many positive responses (combining *strongly agree* and *agree* options) and half as many negative responses (combining *strongly disagree* and *disagree*) during the evaluation than did the CCP document. The larger proportion of *no information to answer* response options in the CCP evaluation results can not completely account for this difference. Further, higher scores were found across most of the dimensions for the CCA document, not only those judged as less relevant by the appraisers. Thus, there is evidence that absolute qualitative differences also played a role. In summary, higher ratings of the CCA document relative to the CCP document can likely be explained by a combination of reporting style and guideline quality factors.

The current evaluation shows that there is correspondence between the two documents regarding common areas of strength and common areas of weakness. Considering

areas of strength, Grilli and his colleagues suggest high quality CPGs from specialty societies should report relevant stakeholder groups, a strategy to identify primary evidence, and a process to grade recommendations based on the strength of studies incorporated into the report.¹⁴ The CCP and CCA documents meet two of these criteria; in both cases, a very elaborate and well thought out system of classifying the recommendations and the type and quality of research used to inform them is outlined. Further, comprehensive memberships of the guideline development groups, particularly in the case of the CCP, and clearly identified target audiences, added credibility to the processes.

In both documents, the reader could typically follow the rationale of why topics were chosen and understand the objectives. The documents were well organized and framed and the recommendations were easily located. Thus *scope and purpose* and *clarity* dimensions were the stronger features of both guideline groups that were well represented.

The third quality component advocated by Grilli was not successfully incorporated into either of the guideline documents.¹⁴ The appraisers agreed that both guideline documents could be improved by being more explicit, particularly in the identification and use of evidence and the linkages between evidence and recommendations. Indeed, the feature that most clearly deviated from an evidence-based perspective of guideline development, was the lack of information describing the processes used to identify and choose the reviewed literature. Indeed, both developers did not explicitly detail the strategy used to search for the evidence, and neither outlined the specific inclusion and exclusion criteria used to select the literature. Thus, for each applicable chapter, the specific body of literature considered, how this literature was chosen, and why other literature was ignored was unclear. Further, there was considerable inconsistency across the chapters, for both groups, with the link between evidence and the recommendations.

An implicit rather than explicit use of evidence not only compromises the judicious consideration of and likely compliance with CPG recommendations, it leads to difficulties in replicating the guideline development process and reconciling differences in recommendations produced on the same topic by the two developers.^{7,9} For example, the recommendations for the CCA *Frequency and Duration of Care* chapter are very different from the CCP coun-

terpart, *Duration of Care for Vertebral Subluxation*. Whereas the former provides very detailed timelines, the latter explicitly indicates that there are no data to substantiate specific time periods for care. Review of the reference lists indicates very little correspondence between the two groups. This difference can not be explained by the different guideline dates (i.e., the more recent CCP document, does not have a reference list that only incorporates literature published after the release of the CCA document). The absence of information describing the literature search and selection process makes it difficult to comment on which one of the two sets of guideline recommendations is more valid. Further, the factors that could explain why such very different recommendations emerged are unknown (e.g., regional differences in care, differences in patient values or clinical culture, etc.).

Finally, there were times when the reports did not establish a link between the evidence and the recommendations, did not indicate the range of clinical options available, and offered ambiguous recommendations that provided little direction for the management of care, features linked to recommendation and uptake. For example, in an observational study exploring factors that facilitate and hinder CPG practice, Grol and his colleagues found almost a two-fold difference in uptake between CPGs with ambiguous recommendations versus clearly worded recommendations (36% vs. 67%, respectively).⁷

The level of editorial independence of the guideline development group from the funders was also unclear in both documents. Lack of editorial independence has the potential to significantly jeopardise the quality of the document. Consider the findings by Barnes and Bero who found a statistically and scientifically meaningful relationship between conclusions of review articles on the effects of passive (“second-hand”) smoking and affiliations with authors.¹³ Here, the investigators found that of the 37% of reviews that concluded no harmful effects of passive smoking, over 70% of these authors had affiliations with the tobacco industry. If the CCA and CCP groups were in fact independent from the funders, an explicit statement to this effect would alleviate the liability inherent in the ambiguity. If there is dependency with funds, this has the capacity to undermine the credibility of the documents.

In summary, the strongest areas for both documents include an identification of guideline objectives and clinical questions/themes, description of target users and patient

populations, a clear guideline structure, credible members on the guideline development groups, and a complete description of the methods used to formulate the recommendations. Areas recommended for improvement in both documents include a more explicit and complete description of the specific bodies of literature under consideration, more detail regarding the systematic methods used to search the evidence, incorporating inclusion and exclusion criteria, statements describing the links between the evidence and recommendations, and greater consistency in creating specific unambiguous recommendations.

Finally, the membership of the research team charged with conducting the evaluation precluded the consideration of clinical content or context. As such, it may be useful to engage in the evaluation process again bringing this perspective to the table. Further, it might be useful to repeat this exercise when updated guideline documents developed by the CCA and CCP are released and when the final version of the AGREE evaluation instrument is available in the public domain.

References

- 1 Woolf SH. Practice guidelines: A new reality in medicine. I. Recent developments. *Arch Intern Med* 1990; 150:1811–1818.
- 2 Henderson D, Chapman-Smith D, Mior S, Vernon H (Eds). *Clinical Practice Guidelines for Chiropractic Practice*. Proceedings of a Consensus Conference Commissioned by the Canadian Chiropractic Association. Supplement to *Journal of the Canadian Chiropractic Association* 1994; 38(1).
- 3 Council on Chiropractic Practice. *Clinical practice guideline: vertebral subluxation in chiropractic practice*. Arizona: Council on Chiropractic Practice, 1998.
- 4 Hayward RSA, Wilson MC, Tunis SR, Bass EB, Guyatt G, for the Evidence-based Medicine Working Group. Users’ guides to the medical literature. Viii. How to use clinical practice guidelines. A. Are the recommendations valid? *JAMA* 1995; 274:570–574.
- 5 The AGREE Collaboration. *Appraisal of Guidelines for Research and Evaluation in Europe (AGREE) Instrument*. May 1999.
- 6 Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *International Journal for Quality in Health Care* 1999; 11:21–28.

- 7 Grol R, Dalhuijzen J, Mookink H, Thomas S, Veld C, Rutten G. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998; 311:237–242.
- 8 Lohr KN. The quality of practice guidelines and the quality of healthcare. In: *Guidelines in health care. Report of a WHO conference. January 1997, Baden-Baden: Nomos Verlagsgesellschaft, 1998.*
- 9 Browman GP. Evidence-Based Cancer Care and Clinical Practice Guidelines. *Am Soc Clin Oncol* 1998; 451–457.
- 10 Brouwers M, Johnston M, Browman G (in press). Results of a prospective study to keep guidelines current. *Proceedings of the ISTAHC 2001 Conference.*
- 11 Moher D, Pham B, Jones A, Cook D, Jadad A, Moher M, Tugwell P, Klass T. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352:609–613.
- 12 Colditz GA, Miller JN, Mosteller F. How study design effects outcomes in comparisons of therapy. I: Medical. *Statistics in Medicine* 1989; 8:441–454.
- 13 Barnes DE, Bero LA. Why review articles on health effects of passive smoking reach different conclusions. *JAMA* 1998; 279:1566–1570.
- 14 Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by speciality societies: the need for critical appraisal. *Lancet* 2000; 355:103–106.
- 15 Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use. Second edition. Oxford: Oxford University Press, 1995.*

HELP SUPPORT CHIROPRACTIC RESEARCH

CANADIAN CHIROPRACTIC RESEARCH FOUNDATION

The C.C.R.F. is a registered Charitable foundation dedicated to quality Chiropractic research. We appreciate your continued support.

Please send your

Tax Deductible Donation

TODAY:

Donations

and/or

Requests for Grant Applications

may be forwarded to:

**Canadian Chiropractic
Research Foundation
1396 Eglinton Avenue West
Toronto, Ontario M6C 2E4**



Dr. Rob Allaby

Treasurer

Canadian Chiropractic Research Foundation