Special Issue

Philadelphia Panel Evidence-Based Clinical Practice Guidelines on Selected Rehabilitation Interventions: Overview and Methodology

Introduction. A structured and rigorous methodology was developed for the formulation of evidence-based clinical practice guidelines (EBCPGs), then was used to develop EBCPGs for selected rehabilitation interventions for the management of low back, neck, knee, and shoulder pain. Methods. Evidence from randomized controlled trials (RCTs) and observational studies was identified and synthesized using methods defined by the Cochrane Collaboration that minimize bias by using a systematic approach to literature search, study selection, data extraction, and data synthesis. Meta-analyses were conducted where possible. The strength of evidence was graded as level I for RCTs or level II for nonrandomized studies. Developing Recommendations. An expert panel was formed by inviting stakeholder professional organizations to nominate a representative. This panel developed a set of criteria for grading the strength of both the evidence and the recommendation. The panel decided that evidence of clinically important benefit (defined as 15% greater relative to a control based on panel expertise and empiric results) in patient-important outcomes was required for a recommendation. Statistical significance was also required but was insufficient alone. Patient-important outcomes were decided by consensus as being pain, function, patient global assessment, quality of life, and return to work, providing that these outcomes were assessed with a scale for which measurement reliability and validity have been established. Validating the Recommendations. A feedback survey questionnaire was sent to 324 practitioners from 6 professional organizations. The response rate was 51%. Results. Eight positive recommendations of clinical benefit were developed. These recommendations were mainly in agreement with previous EBCPGs, although some were not covered by other EBCPGs. There was wide agreement with these recommendations from practitioners (greater than 75% agreement). For several interventions and indications (eg, thermotherapy, therapeutic ultrasound, massage, electrical stimulation, mechanical traction), there was a lack of evidence regarding efficacy. Conclusions. This methodology of developing EBCPGs provides a structured approach to assessing the literature and developing EBCPGs that incorporates clinicians' feedback and is widely acceptable to practicing clinicians. Further well-designed RCTs are warranted regarding the use of several interventions where evidence was insufficient to make recommendations. [Philadelphia Panel Evidence-Based Clinical Practice Guidelines on Selected Rehabilitation Interventions: Overview and Methodology. *Phys Ther.* 2001;81:1629–1640.]

Bone and Joint DECADE AMERICA 2000-2010

APTA is a sponsor of the Decade, an international, multidisciplinary initiative to improve health-related quality of life for people with musculoskeletal disorders.

Key Words: *Clinical practice guidelines, Evidence-based practice, Meta-analysis, Musculoskeletal conditions, Physical therapy, Practitioner feedback survey, Rehabilitation, Systematic review.*

INTRODUCTION

There is an urgent need for rehabilitation specialists to demonstrate the efficacy of their interventions.[1] The most recognized way to fulfill this goal is to use scientific evidence to guide health care professionals' practice.[2–5] Professional associations and colleges have also encouraged rehabilitation specialists to use the best-available clinical research to guide their practice.[6,7] The development and use of critical appraisal tools has been promoted to help rehabilitation specialists adopt evidence-based practice.[8,9]

Evidence-based practice tools include evidence-based clinical practice guidelines (EBCPGs), validated outcome measures, education, and continuing professional development. *Evidence-based clinical practice guidelines* have been defined as systematically developed statements to help practitioners and clients with decisions about appropriate health care for specific clinical circumstances.[10,11] Evidence-based clinical practice guidelines are a rapidly emerging technology with considerable potential to alter the process of clinical decision making in fundamental ways. Furthermore, the appropriate use of EBCPGs has been demonstrated to improve both the process of care and client health outcomes.[12]

Recently, there has been much enthusiasm for the establishment of EBCPGs to assist clinical decision making and to improve health outcomes.[6,7,13,14] The development of EBCPGs involves 5 steps: defining the question, collecting the evidence, synthesizing the results, making a recommendation based on the results, and grading the strength of the recommendation.[15]

The management of musculoskeletal pain is complex and involves different types of practitioners. Rehabilitation specialists managing clients with musculoskeletal problems can use multidisciplinary EBCPGs, such as:

Philadelphia Panel Members:

Clinical Specialty Experts:

John Albright, MD (Orthopaedic Surgeon), American Academy of Orthopaedic Surgeons, USA
Richard Allman, MD (Internist, Rheumatologist), American College of Physicians, USA
Richard Paul Bonfiglio, MD (Physiatrist)
Alicia Conill, MD (Internist), University of Pennsylvania, Philadelphia, Pa, USA
Bruce Dobkin, MD (Neurologist), American Academy of Neurology, USA
Andrew A Guccione, PT, PhD (Physical Therapist), American Physical Therapy Association, USA
Scott Hasson, PT, EdD (Physical Therapist), American College of Rheumatology, Association of Health Professionals, USA
Randolph Russo, MD (Physiatrist), American Academy of Physical Medicine and Rehabilitation, USA
Paul Shekelle, PhD (Internist), Cochrane Back Group
Jeffrey L Susman, MD (Family Practice), American Academy of Family Physicians, USA

Ottawa Methods Group:

George A Wells, PhD[1,2,3,5]
Peter Tugwell, MD, MSc[1,2,3,5]
Lucie Brosseau, PhD[1,2,4,6]
Vivian A Robinson, MSc[1, 5]
Ian D Graham, PhD[1,2,3]
Beverley J Shea, RN, MSc[1,3,5]
Jessie McGowan, MLIS, Director, Medical Library[3]

1. Clinical Epidemiology Program Unit, Ottawa Health Research Institute, Ottawa Hospital, Civic Campus, Ottawa, Ontario, Canada
2. Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada
3. Department of Medicine, Ottawa Hospital, General Campus, Ottawa, Ontario, Canada
4. Program of Physiotherapy, School of Rehabilitation Sciences, Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada
5. Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada
6. Ministry of Ontario Health (Canada)

Address all correspondence and requests for reprints to: Peter Tugwell, MD, MSc, Chair, Centre for Global Health, Institute of Population Health, 1 Stewart St, Rm 312, Ottawa, Ontario, Canada K1N 6N5 (ptugwell@uottawa.ca).

This study was financially supported by an unrestricted educational grant from the Cigna Foundation, Philadelphia, Pa, USA, the Ministry of Human Resources and Development, Government of Canada (Summer Students Program), and the Ontario Ministry of Health and Long-Term Care (Canada).

Ian Graham is a Medical Research Council Scholar, Canadian Institutes of Health Research (Canada).

Acknowledgments: Summer students: Sarah Milne, Michael Saginur, Marie-Josée Noël, Mélanie Brophy, Anne Mailhot.

(1) the guidelines of the Québec Task Force on Spinal Disorders,[16] (2) the American Health Care and Policy Research (AHCPR) guidelines for acute low back pain,[17,18] (3) the *British Medical Journal Clinical Evidence*,[19] and (4) the American College of Rheumatology (ACR) guidelines for knee osteoarthritis.[20] Some limitations of these existing EBCPGs are: (1) the clinical practice areas were limited, (2) the EBCPGs were not based on updated systematic reviews, or systematic reviews were not conducted by the EBCPGs' developers, (3) the recommendations are too broad or are not specific enough, (4) EBCPGs are not recommended related to a specific outcome or a specific sample of rehabilitation clients (absence of inclusion/exclusion criteria), (5) the EBCPGs do not distinguish between acute and chronic conditions, (6) the EBCPGs do not provide a clear definition of the intervention, (7) EBCPGs are often based on comparative studies or nonplacebo comparisons, and (8) the EBCPGs do not use a recognized classification to grade the strength of the clinical recommendations.[21] In the past 5 years, a growing interest in EBCPGs has developed among North American occupational therapists and physical therapists.[6,7,14,22] More specifically, Canadian physical therapists developed their own EBCPGs on spinal manipulations[22] and on suctioning.[23]

Rehabilitation interventions for neck, low back, shoulder, and knee pain are a high-volume activity, and EBCPGs based on The Cochrane Collaboration reviews, do not exist. The aim of this article is to describe the methodology used to develop evidence-based recommendations for rehabilitation interventions for musculoskeletal pain in 4 areas: shoulder, knee, low back, and neck. The aim of the developing the EBCPGs was to improve appropriate use of rehabilitation interventions. The target users of these EBCPGs are physical therapists, physiatrists, orthopedic surgeons, rheumatologists, family physicians, and neurologists.

Formation of Panel

The Ottawa Methods Group initiated the formation of the panel by soliciting nominations of clinical specialty experts of the Philadelphia Panel from professional organizations that are interested in the care of patients with neck, back, knee, and shoulder pain. The organizations were asked to consider nominating panelists with: (1) clinical expertise in the management of musculoskeletal pain and (2) familiarity with EBCPGs.

Nine individuals were nominated as clinical specialty experts of the Philadelphia Panel, representing the fields of family medicine, internal medicine, neurology, orthopedic surgery, physical medicine and rehabilitation, physical therapy, rheumatology, and spine research. The Philadelphia Panel includes all members of the Ottawa Methods Group and the 9 clinical specialty experts.

The panel chair (PT) formed a research and support staff that included individuals with expertise in rehabilitation interventions, methodology, meta-analyses, and development and assessment of EBCPGs. The staff screened articles and constructed evidence tables for articles according to the methods described below. These evidence tables were presented to the clinical specialty experts of the Philadelphia Panel for review and interpretation. The panel used these tables as the basis for developing recommendations.

Identifying and Refining the Subject Area

The clinical questions were defined by discussions with potential users of the EBCPGs. The rehabilitation interventions and clinical conditions were limited to high-volume activities for which no current EBCPGs exist.

An explicit set of selection criteria were established *a priori* in order to appropriately address the topics of interest and minimize the time spent reviewing irrelevant material. The selection criteria were reviewed by the Ottawa Methods Group. The selection criteria were laid out explicitly with a checklist format that ensured a systematic and reproducible approach to study selection. Briefly, studies were included if they met the following criteria:

Population: Outpatients with shoulder, neck, low back, or neck pain were included. Patients with scoliosis, cancer, or pulmonary, neurologic (except peripheral nerve injuries), pediatric, cardiac, dermatologic, psychiatric, or multiple conditions were excluded. Individuals with no known pathology or impairments also were excluded.

Interventions: Interventions were selected for review based on a perceived need for EBCPGs due to frequent use and insufficient evidence, based on clinical experience. Therapeutic exercise, massage, transcutaneous electrical nerve stimulation (TENS), thermotherapy, ultrasound, electrical stimulation, and combinations of these therapies were included. The following interventions were excluded due to either a sufficient body of knowledge or less frequent use: manipulation, manual therapy, swimming pool exercise, behavioral, educational, functional restoration, and psychosocial interventions. Iontophoresis was excluded because it includes a mix of medication and therapeutic ultrasound, and medication is not a physical rehabilitation intervention. Patients given educational pamphlets were accepted as a central group, but educational interventions that included instruction (either group or individual) by videotape or by a therapist/educator were excluded. Surgery,

electroanalgesia, and inpatient interventions (eg, cryocuff or continuous passive motion provided in hospital) were excluded.

Acceptable comparison interventions were placebo, no treatment, or one of the interventions of interest. Concurrent therapy (eg, exercise, pamphlets) was accepted if it was provided to both treated and comparison groups equally.

Trial Designs: All comparative controlled studies were included such as randomized controlled trials (RCTs), controlled clinical trials, cohort studies, and case-control studies. Case series and uncontrolled cohort studies were excluded.

Outcomes: Outcomes were selected based on their clinical relevance. Studies were included if they measured one of the following outcomes: pain, function, strength, range of motion, return to work, patient satisfaction, activities of daily living, or quality of life (QOL). Psychological outcomes (eg, depression) and physiological outcomes (eg, skin temperature, biochemical markers) were excluded. Cardiopulmonary function and postural assessment also were excluded.

Identifying and Assessing the Evidence

To answer the clinical questions, systematic reviews were performed for all rehabilitation interventions of interest and the 4 clinical conditions, according to the methods of The Cochrane Collaboration.[24]

Before reviews were conducted *de novo*, the Cochrane Database of Systematic Reviews was searched for existing Cochrane reviews of the interventions and conditions of interest. Several existing Cochrane reviews addressed the interventions and clinical conditions of interest, but did not answer the clinical questions because those reviews looked at different interventions,[25] were restricted to double-blind trials,[26] excluded relevant studies,[27] or used different outcomes and analytic techniques.[28]

Identifying the Evidence: A literature search was conducted according to the Cochrane methodology for the identification of RCTs, modified to identify controlled clinical trials, cohort studies, and case-control studies.[29,30] The electronic search strategy was designed based on the defined clinical questions specifying the populations, interventions, outcomes, and study designs that were of interest. Electronic searches were conducted up to July 1, 2000, in MEDLINE from 1962, EMBASE from 1988, CINAHL from 1982, the Cochrane Controlled Trials Register, HEALTHSTAR from 1975, the database of the Cochrane Field of Rehabilitation and Related Therapies (based in Denmark), and PEDro (Physiotherapy Evidence Database 2000 update). Reference lists of included studies and other meta-analyses were hand-searched for relevant articles. The members of the Philadelphia Panel (experts from rheumatology, orthopedic surgery, neurology, physical therapy, physiatry, back pain and internal medicine, and family medicine) were asked whether any additional studies had been missed.

Assessing the Evidence: The relevance of studies retrieved using electronic searching was assessed by 2 independent reviewers who screened the titles and abstracts, using the predetermined checklist of selection criteria. The systematic reviews were restricted to articles published in English, French, or Spanish. Any article identified by one reviewer as potentially relevant was retrieved for closer review. Upon retrieval of the full article, 2 independent reviewers determined relevance to the clinical questions.

Summarizing the Evidence

Data were extracted by 2 independent reviewers from the included studies, using predetermined paper-based forms. These forms collected data regarding the benefits and harms of the intervention as well as population characteristics, trial design, allocation concealment, and details of the interventions. These reviewers also assessed methodological quality of randomization, double-blinding, and description of withdrawals and dropouts using a validated scale (Appendix).[31,32] Differences in data extraction or quality assessment were resolved by consultation with a third reviewer.

Synthesizing the Evidence

The number of included studies was presented graphically in a 3-axis "cityscape" (Fig. 1), where each clinical condition was represented by a "street" of rehabilitation interventions, the height of which represented the number of studies identified for that clinical condition and intervention. This schematic was used to prioritize the analysis of data.

CLINICAL RELEVANCE

The results were presented in tables with 2 shaded columns showing the absolute benefit and the relative difference in the change from baseline. Absolute benefit was calculated as the improvement in the treatment group less the improvement in the control group, in the original units (Tab. 1). Relative difference in the change from baseline was calculated as the absolute benefit divided by the baseline mean (weighted for the treatment and control groups). The relative difference in change was used to provide clinically meaningful information about expected improvement relative to the placebo or untreated group with each intervention. For this analysis, results from individual trials were not combined statistically. Rather, results from individual

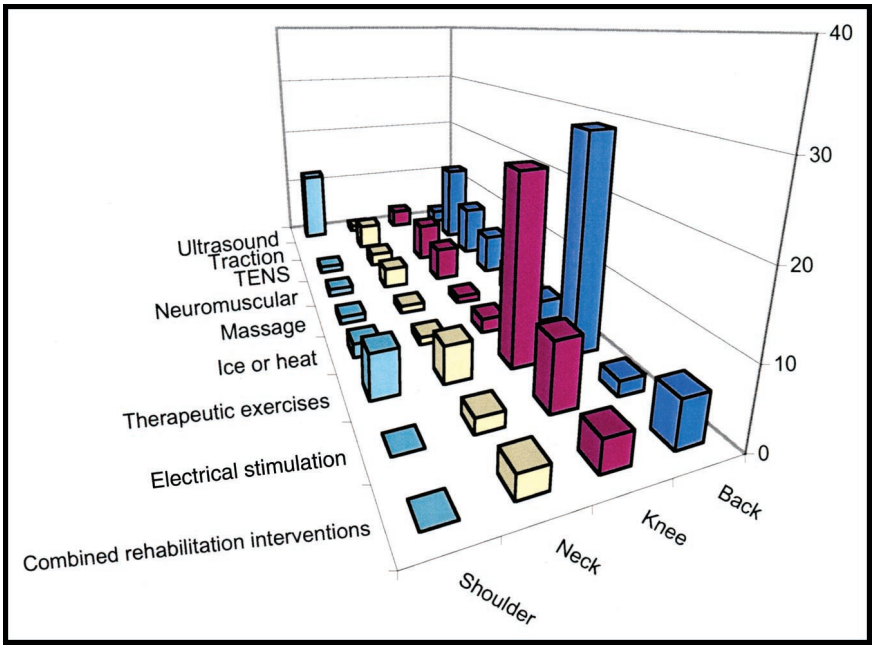

Figure 1.
Cityscape, showing number of included trials for each intervention for neck, back, shoulder, and knee pain. TENS=transcutaneous electrical nerve stimulation.

trials were presented in a table, allowing the comparison of the percentage of improvement in each trial. An example of such a table is presented in Table 1 for therapeutic exercise for subacute low back pain.

STATISTICAL SIGNIFICANCE

Meta-analysis was used to analyze the difference between treatment and control groups at the end of study.[24] For continuous outcomes, results were analyzed as weighted mean differences, where the weighting factor was determined by the inverse of the variance. Where the same concept was measured with different scales (eg, pain), standardized mean differences were used to combine end-of-study results. For dichotomous outcomes, relative risks were calculated. Heterogeneity was tested with Cochrane's Q test. Fixed-effects models were used throughout, unless heterogeneity was significant ($P<.05$), in which case random effects models were considered.

The pooled results were presented in a graphical format, using the Review Manager (RevMan) computer program, Version 4.1 for Windows,* showing the point estimate (difference between treatment and control groups) and the 95% confidence intervals for each trial and for the pooled estimate (Fig. 2).

* Oxford, England: The Cochrane Collaboration, 2000.

CATEGORIZING THE EVIDENCE

In order to select a method of categorizing the evidence according to susceptibility to bias, a literature search was conducted for existing methods of grading the evidence for the development of evidence-based recommendations.

The grading systems of the Canadian Task Force on Periodic Health Examination (CTFPHE),[33] Cancer Care Ontario,[34] the AHCPR guidelines for acute low back pain,[17] and the guidelines of the Quebec Task Force on Spinal Disorders[16] and the grading system recommended by Guyatt et al[15,35] were reviewed. Of these, the grading system used by the CTFPHE was selected because of its ability to grade both the direction of results and the strength of the trial designs.

The CTFPHE system was modified because the limited data available made it questionable whether the negative classifications (D and E) could be justified (Tab. 2). The results of the systematic reviews were categorized according to this modified system by the Ottawa research team and summarized in a master grid (Tab. 3).

TRANSLATING EVIDENCE INTO A CLINICAL PRACTICE GUIDELINE

The results of the evidence synthesis were sent to the Philadelphia Panel for their review. A 1-day panel face-to-face meeting was used to determine how to incorporate opinion into the interpretation of results as well as how to apply this methodology.

USING AND GATHERING OPINION

At the panel meeting, 4 hours were spent on defining a transparent and reproducible method of assessing the evidence synthesis and making recommendations, with the consensus of all panelists.

Outcomes

The panel reviewed the relevance of key outcomes for deciding whether a given intervention has clinical benefit. The panel decided to take the clinician and patient perspective rather than a payer perspective. The following outcomes were agreed upon as having clinical importance:

1. Pain

2. Function/QOL

Table 1.
Example of Clinical Relevance: Therapeutic Exercises for Chronic Low Back Pain (>12 Weeks): Pain at 1 Month[a]

Study	Treatment Group	Outcome (Scale)	No. of Patients	Baseline Mean	End of Study Mean	Absolute Benefit	Relative Difference in Change From Baseline
Frost,[b] 1995	E: strength, stretch, aerobic exercises	Sensory Pain (0–100 VAS)	36	20.9	12.1	−5.30 (I) on 100 mm VAS	−23% (I)
	C: control		35	25.6	22.1		
Deyo et al,[45] 1990	E: stretch	Pain improvement (0–100)	63	NA	47.9	+7 (I) on 100-point scale	+7% (I)
	C: control		63	NA	40.9		
Spratt,[c] 1993	E: McKenzie	Pain (0–10 VAS)	21	5.6	6.85	−1.12 (I) on 10-cm VAS	−20% (I)
	C: control		17	5.84	5.97		
Hansen,[d] 1993	E: strength exercises	Pain (0–9 VAS)	44	5.0	4.1	−3.00 (I) on 9-cm VAS	−60% (I)
	C: control		28	5.0	7.1		
Risch,[e] 1993	E: strength, stretching exercises	Pain (West-Haven Yale, 0–12)	31	3.4	2.9	−0.90 (I) on 12-point scale	−26% (I)
	C: control		23	3.7	4.1		

[a] (I) indicates improvement better in treatment group than in control group, negative or positive depending on anchors for the scales. VAS=visual analog scale, NA=not available.
[b] Spratt KF, Weinstein JN, Lehmann TR, et al. Efficacy of flexion and extension treatments incorporating braces for low-back pain patients with retrodisplacement, spondylolisthesis, or normal sagittal translation. *Spine.* 1993;18:1839–1849.
[c] Frost H, Klaber M, Moser JS, Fairbank JC. Randomised controlled trial for evaluation of fitness programme for patients with chronic low back pain. *BMJ.* 1995;310(6973):151–154.
[d] Risch SV, Norvell NK, Pollock ML, et al. Lumbar strengthening in chronic low back pain patients: physiologic and psychological benefits. *Spine.* 1993;18:232–238.
[e] Hansen FR, Bendix T, Skov P, et al. Intensive, dynamic back-muscle exercises, conventional physiotherapy, or placebo-control treatment of low-back pain: a randomized, observer-blind trial. *Spine.* 1993;18:98–108.

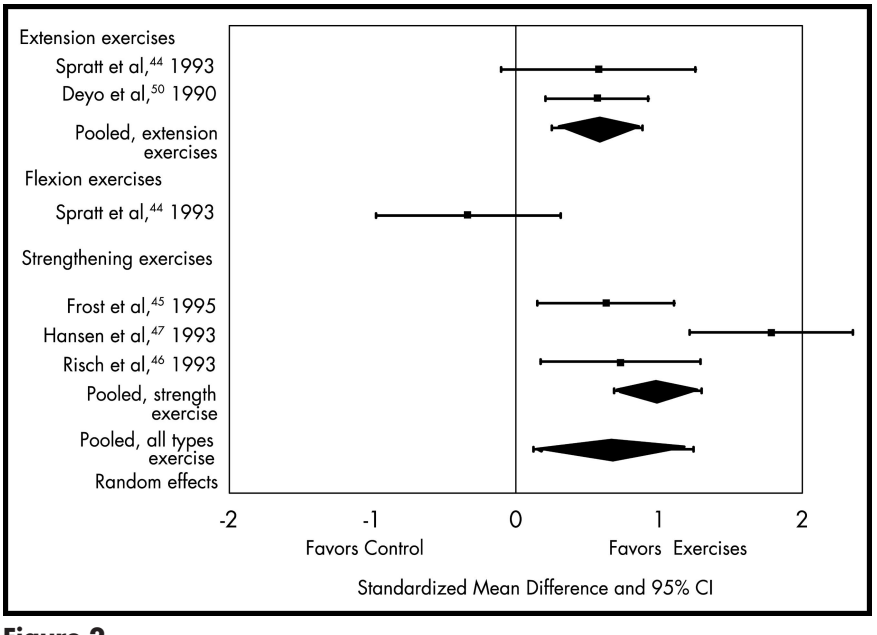

Figure 2.
Example of Review Manager analysis. CI=confidence interval. See Table 1 footnotes for references for studies cited.

Table 2.
Modified Canadian Task Force on Periodic Health Examination Grading System

Grading of Evidence	
I	Evidence from at least 1 properly randomized controlled trial (RCT)
II-1	Evidence from well-designed controlled trials without randomization
II-2	Evidence from well-designed cohort or case-control analytic studies, preferably from more than 1 center or research group
II-3	Evidence from comparisons between times or places with or without the intervention. Dramatic results in uncontrolled experiments could also be included here.
III	Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees
Grading of Recommendations	
A	Good evidence to support the recommendation that the intervention be specifically considered
B	Fair evidence to support the recommendation that the intervention be specifically considered
C	Poor evidence regarding inclusion or exclusion of a intervention, but recommendations may be made on other grounds

Table 3.
Master Grid for Back Pain Prior to Philadelphia Panel Review[a]

	Acute (1–6)	**Chronic (7–15)**	**Post-surgery (16)**	**Spondylolysis (17)**
Traction	√ C, I	√ C, I	nd	nd
Exercise	√ B, I	√ A, I	√ A, I	nd
Massage	√ C, II	√ C, I	nd	nd
Thermotherapy	nd	√ C, II	nd	nd
Therapeutic ultrasound	√ C, II	√ C, II	nd	nd
TENS	√ C, I	√ C, I	nd	nd
Electrical stimulation	nd	√ C, I	nd	nd
Neuromuscular re-education	√ C, I	√ C, II	nd	√ C, I
Combined physical therapy	nd	√ C, II	nd	nd

[a] TENS=transcutaneous electrical nerve stimulation. √ indicates that a recommendation could be formed, based on evidence from trials. nd=no data. Grades of evidence: I=randomized controlled trial, II=controlled clinical trial without randomization. Grades of recommendation: A=good, B=fair, C=poor evidence to recommend.

3. Return to work

4. Patient global assessment (patient's assessment of overall disease activity or improvement)

5. Patient satisfaction

The panel believed that scales demonstrated to be valid and responsive to change should be required to support a positive recommendation (A or B). Other outcomes, although providing useful information in studies, were believed to be insufficient to warrant a grade A or B recommendation.

Clinical Importance and Statistical Significance

There is some empirical evidence in rheumatology that greater than 20% improvement is viewed by patients as a clinically important difference between 2 interventions and that this discriminates active from placebo/control in all the RCTs reviewed for the ACR.[36] The ACR criterion of 20% improvement was developed in 3 steps: (1) a survey of rheumatologists using patient scenarios to identify the cutoff that corresponds best with rheumatologists' impression of improvement, (2) testing, in existing data sets, which cutoff criteria maximally discriminated effective from placebo and minimized the placebo response, and (3) testing of the 8 remaining cutoff definitions for ease of use and best accordance with clinician impression of improvement.

A difference of 2 points on the Roland scale (0–24 scale) is widely used as a minimally important change for back pain, and this amounts to approximately 15% improvement relative to the control group (when considering the usual baseline Roland scale score of 11 or 12).[37]

The panel decided to accept 15% difference between groups as clinically important and that a 15% or greater difference and statistical significance were required for grade A and B recommendations. The panel decided that a C+ recommendation could be used to demonstrate that a potential clinically important benefit of 15% or greater was found but without statistical significance.

Defined Diagnosis and Reproducible Study Population

For any recommendation, the panel decided that the diagnosis and population must be described in sufficient detail to be of use clinically. Furthermore, the panel decided that studies that combined clinically heteroge-

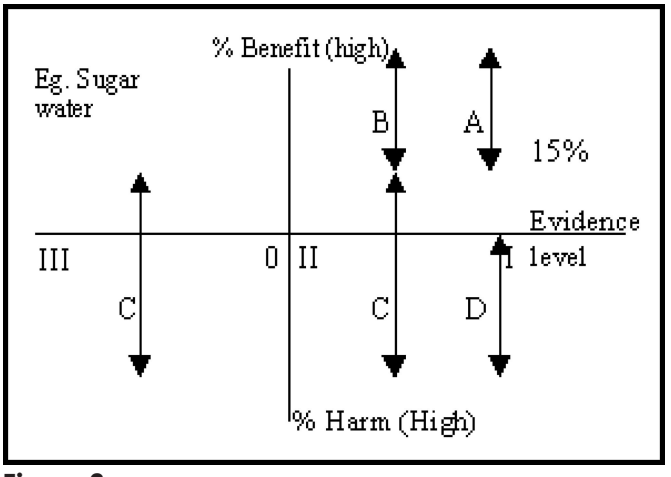

Figure 3.
Classification system agreed to by Philadelphia Panel. For grade C, statistical significance is unimportant (ie, clinical importance is not met; therefore, statistical significance is irrelevant). RCT=randomized controlled trial, CCT=nonrandomized controlled clinical trial. See Table 4 for details of the classification system.

neous populations should be excluded (eg, patients with acute and chronic low back pain in the same trial).

Study Design and Methodologic Quality

The panel decided that evidence from one or more RCTs of a clinically important benefit (≥15%) that is statistically significant was necessary for a grade A recommendation. A grade B recommendation would be given for a clinically important benefit (≥15%) that is statistically significant if the evidence was from observational studies or controlled clinical trials. Because there is less confidence in the results from nonrandomized trials, controlled clinical trials were accepted only if they scored 3 or more out of 5 on the Jadad scale, which gives 2 points for randomization, 2 points for blinding, and 1 point for describing withdrawals. Evidence of clinical importance (≥15%) but not statistical significance would be considered a grade C+ recommendation. Based on these decisions, grade C recommendations would be given to those interventions where an appropriate outcome was measured in a study that met the inclusion criteria and no clinical importance was shown. This grading system is conceptualized in Figure 3 and described in detail in Table 4.

No recommendation was possible when the data were insufficient, and these EBCPGs were assigned a classification of "Insufficient Data" (ID). This classification was used because there were (1) interventions where no relevant outcome using a validated scale was reported, (2) studies with ≤10 patients randomly assigned to the trial, and (3) interventions where only head-to-head trials were available.

Grading Recommendations

Once the methodology of gathering opinion and interpreting the evidence was defined, the grading of recommendations proceeded quickly. Fifty-two clinical questions were addressed in 3 hours. The revised grading was summarized using the same master grid approach, which allowed comparison with the earlier recommendations (Tab. 5). Each positive recommendation was summarized as a one-page guideline.

External Review of the EBCPGs

External review by practitioners and incorporation of their comments into the EBCPGs are important to ensure the uptake and relevance of guidelines.[34] The guidelines were sent to the Philadelphia Panel for review. In order to judge the clinical usefulness, the 9 positive recommendations were sent to 324 practitioners for their feedback. Practitioners were selected from membership lists of key professional associations, including physical therapists, orthopedic surgeons, physiatrists, back specialists, family practitioners, and rheumatologists. Practitioners were asked 3 questions for each guideline (Tab. 6). This feedback was then discussed by the panel, and the guidelines were revised accordingly. In this way, the feedback from the practitioners was incorporated into the completed EBCPGs.[38]

Reporting

The process of reporting the evidence of the results of the systematic reviews conducted throughout this project follow the recommendations made by QUOROM.[39] Each of the systematic reviews will be published separately in both paper-based format and as Cochrane reviews in cases where there is not one currently existing in The Cochrane Library.

Evaluation of the EBCPGs

These EBCPGs were assessed using a quality assessment tool called AGREE.[21] This tool consists of 6 dimensions measured on a 4-point scale (where 1 represents "strongly agree" and 4 represents "strongly disagree"). The overall scores were reached by consensus among 5 independent rehabilitation specialists. The results for the dimensions were: 12/12 for scope and purpose, 7/8 for stakeholder involvement, 18/24 for rigor of development, 11/12 for clarity and presentation, 9/24 for applicability, and 8/8 for editorial independence.

DISCUSSION

The Philadelphia Panel has designed a consensus-based and rigorous methodology in order to develop EBCPGs, using a transdisciplinary approach. Eight EBCPGs were defined where rehabilitation interventions were shown to have a beneficial effect on patient-important outcomes. However, these EBCPGs are subject to a number of methodological limitations, as with all such reviews.

Table 4.
Details of Philadelphia Panel Classification System

	Clinical Importance	Statistical Significance	Study Design[a]
Grade A	>15%	*P*<.05	RCT (single or meta-analysis)
Grade B	>15%	*P*<.05	CCT or observational (single or meta-analysis), with a quality score of 3 or more on the 5-point Jadad methodologic quality checklist
Grade C+	>15%	Not significant	RCT or CCT or observational (single or meta-analysis)
Grade C	<15%	Unimportant[b]	Any study design
Grade D	<0% (favors control)		Well-designed RCT with >100 patients

[a] RCT=randomized controlled trial, CCT=controlled clinical trial.
[b] For grade C, statistical significance is unimportant (ie, clinical importance is not met; therefore, statistical significance is irrelevant).

Table 5.
Master Grid for Back Pain After Philadelphia Panel Review[a]

	Acute	Subacute LBP	Chronic	Post-surgery
Traction	√ C	√ C	√ C	nd
Exercise	√ A	√ A	√ A	√ A
Massage	√ ID	nd	√ ID	nd
Thermotherapy	nd	nd	√ ID	nd
Therapeutic ultrasound	√ C	nd	√ C	nd
TENS	√ C	nd	√ C	nd
Electrical stimulation	nd	nd	√ ID	nd
Neuromuscular re-education	√ ID	nd	√ ID	nd
Combined physical therapy	nd	nd	√ C	nd

[a] LBP=low back pain, TENS=transcutaneous electrical nerve stimulation. √ indicates a recommendation could be made based on evidence that met the criteria for outcomes, study designs of the Philadelphia Panel. Grade A=good evidence to include intervention (>15% relative improvement, statistical significance, from randomized controlled trials), grade B=fair evidence to include intervention (>15% relative improvement, statistical significance, from nonrandomized controlled trials), grade C=poor evidence to include or exclude intervention (<15% relative improvement). nd=no data. ID=insufficient data; data were identified from clinical trials, but did not meet criteria because they: (1) had no placebo or untreated group or (2) did not measure a validated, clinically important outcome.

The Philadelphia Panel decided, based on experience, that the outcomes of primary clinical importance are: pain, functional status, patient global assessment, QOL, return to work, and patient satisfaction. All outcome measurements were required to have demonstrated validity and reliability. However, these outcomes were assessed using multiple scales and methods. Standardized measurement of outcomes is needed to facilitate scientific advances in the efficacy of rehabilitation interventions.[40,41] More information on responsiveness and the minimally important change are needed for many of the instruments used to measure the effects of rehabilitation interventions.[37,41–43] Validity, reliability, and sensitivity to change of outcome measurements should be evaluated in developing a core set of standardized outcome measures.[44]

Methodologic quality of the included trials rarely reached 4 or greater out of 5 on the Jadad scale (Appendix). Randomization was rarely fully adequate (ie, performed using computerized random number lists). Insufficient information about the treatment assignment procedure was noted in several RCTs. Complete blinding is difficult to achieve for most rehabilitation interventions because of visual and other sensory differences between treatment and placebo as well as unintended communication between patient and evaluator.[45] Few investigators reported adequate information regarding withdrawals and loss to follow-up or indicated whether they were considered in the data analysis. These weaknesses contribute to the lower-quality assessment scores in many of the systematic reviews conducted on interventions of interest.

Various methodological biases could have been introduced in the individual trials. A misclassification bias related to the condition studied is present with the lack of precise medical and physical therapist diagnoses observed.[40,43,46–49] Selection bias may have occurred with the presence of heterogeneity of clinical characteristics such as age, prevalent versus incident cases, stages of the disease, level of pain, and presence or absence of neurological deficits. However, differences in disease duration were minimized in these guidelines by excluding studies with a mix of acute and chronic conditions or mixed diagnoses other than sciatalgia. Characteristics of the device parameters and of the therapeutic application[50] could also make a difference in the effect size. Publication bias, where only the trials with positive findings have been published, may cause an exaggeration of the treatment effect.[51] The effect of publication bias could not be assessed due to the few number of trials in each meta-analysis. A language bias was introduced because the Philadelphia Panel reviewed

Table 6.
Practitioner Feedback Survey Questions

1. Do you agree with this evidence-based recommendation?	___ Agree	___ Disagree	
2. Will a majority of your colleagues agree with the recommendation?	___ Yes		___ No
3. Will you follow this recommendation?	___ Yes	___ Already do	___ No

only studies published in English, French, or Spanish, although recent studies have shown no difference between those including only English and those including other languages.[39]

The Philadelphia Panel agreed that clinical importance be defined as a clinical improvement of 15% or more relative to control. Grade A or B recommendations were required to demonstrate both clinical importance and statistical significance. Although the use of clinical importance is an important step in making the results of systematic reviews more relevant to clinicians, this definition was arrived at using evidence from rheumatology and back pain research and may not be applicable to all rehabilitation interventions and outcomes. Furthermore, the systematic reviews used for these guidelines may have low statistical power in detecting minimally clinical important differences due to small sample sizes and few studies.[52] These issues contribute to nonconclusive results for several interventions.

The therapeutic application of most rehabilitation interventions is based on empirical experience.[2,53,54] Often, multiple rehabilitation interventions are used in one treatment session, depending on various factors such as disease duration, stage of the condition, and the effects of prior treatments. The measurement of patient-important effects is complex.[40,55] The practice of rehabilitation requires a better theoretical basis[40,50,56,57] and well-designed, controlled research.[58]

Ensuring that the EBCPGs are sound before recommending their use is essential to policymakers responsible for guideline programs, and a formal appraisal should be an integral part of those programs.[21] Various groups have developed methods of assessing the quality of EBCPGs.[59] The evaluation of the Philadelphia Panel EBCPGs using AGREE yielded a very high score for dimensions 1, 2, 4, and 6 (purpose, stakeholder involvement, clarity, and editorial independence, respectively), with lower scores for dimensions 3 (rigor of development) and 5 (applicability). The rigor of development was low due to poor reporting of the side effects and risks. Side effects and risks were not reported in the primary trials and therefore were not included in the EBCPGs. The applicability was low, particularly in terms of identifying potential organizational barriers, cost implications, and methods of applying and monitoring the EBCPGs. The Ottawa Methods Group is planning implementation studies with the EBCPGs after publication, which is one of the requirements identified in the literature for uptake of EBCPGs.[60]

CONCLUSION

This methodology for developing EBCPGs is based on systematic reviews and meta-analysis as well as expert opinion. The evidence was assessed with a standardized approach that emphasized clinical importance rather than statistical significance.

References

1 Stratford PW. Who will decide the efficacy of physiotherapeutic interventions? *Physiotherapy Canada.* 1999;51:235–238.

2 Delitto A. Clinicians and researchers who treat and study patients with low back pain: are you listening? *Phys Ther.* 1998;78:705–707.

3 McDonald J, Overhage JM. Guidelines you can follow and can trust: an ideal and an example. *JAMA.* 1994;271:872–873.

4 Sackett DL, Straus SE, Richardson WS, et al. *Evidence-Based Medicine: How to Practice and Teach EBM.* Edinburg, Scotland: Churchill Livingstone; 2000.

5 Helewa A, Walker J. *Critical Evaluation of Research in Physical Rehabilitation.* Philadelphia, Pa: WB Saunders Co; 2000.

6 *Guide to Physical Therapist Practice: Second Edition.* Alexandria, Va: American Physical Therapy Association; 2001.

7 *Clinical Practice Guidelines: A Discussion Paper for the Canadian Physiotherapy Association.* Toronto, Ontario, Canada: Canadian Physiotherapy Association; 1996

8 Balmer S, King J. Linking clinical practice and evidence. *Physiotherapy Canada.* 1998;50:181–183.

9 Bohannon RW, LeVeau BF. Clinicians' use of research findings: a review of literature with implications for physical therapists. *Phys Ther.* 1986;66:45–50.

10 What are "clinical practice guidelines"? In: Rothstein JM, Delitto A, Scalzitti DA. Understanding AHCPR Clinical Practice Guideline No. 14: Acute Low Back Problems in Adults. *PT Magazine.* 1995;3(9 suppl):3–5.

11 Woolf SH. Practice guidelines: a new reality in medicine, I: recent developments. *Arch Intern Med.* 1990;150:1811–1818.

12 Grimshaw JM, Freemantle N, Wallace S, et al. Developing and implementing clinical practice guidelines. *Qual Health Care.* 1995;4:55–64.

13 Canadian Association of Occupational Therapists. *Guidelines for the Client-Centered Practice of Occupational Therapy.* Ottawa, Ontario, Canada: Health and Welfare (Canada); 1991.

14 Woods-Duncan P. Synthesis of intervention trials to improve motor recovery following stroke. *Topics in Stroke Rehabilitation.* 1997;3:1–20.

15 Browman GP, Levine MN, Mohide EA, et al. The practice guidelines development cycle: a conceptual tool for practice guidelines development and implementation. *J Clin Oncol.* 1995;13:502–512.

16 Québec Task Force on Spinal Disorders. Scientific approach to the assessment and management of activity-related spinal disorders: a monograph for clinicians. *Spine.* 1987;12:51–59.

17 *Acute Low Back Problems in Adults.* Clinical Practice Guideline No. 14. Rockville, Md: Agency for Health Care Policy and Research, US Department of Health and Human Services; 1994. AHCPR Publication No. 95–0642.

18 Rothstein JM, Delitto A, Scalzitti DA. Understanding AHCPR Clinical Practice Guideline No. 14: Acute Low Back Problems in Adults. *PT Magazine.* 1995;3(9 suppl):1–24.

19 *Clinical Evidence: A Compendium of the Best Available Evidence for Effective Health Care.* London, England: BMJ Publishing Group; 1999(2).

20 Hochberg MC, Altman RD, Brandt KD. Guidelines for the medical management of osteoarthritis, part II: osteoarthritis of the knee. *Arthritis Rheum.* 1995;38:1541–1546.

21 Cluzeau F. *Appraisal Instrument of Guidelines for Research and Evaluation (AGREE) Instrument.* Available at: www.agreecollaboration.org.

22 *Clinical Guidelines on Spinal Manipulation.* Ontario, Canada: College of Physiotherapists of Ontario; 1998.

23 Brooks D, Solway S, Graham I, et al. *Interdisciplinary Suctioning Clinical Practice Guidelines Initiative: Technical Report.* Ontario, Canada: Colleges of Physiotherapists of Ontario, the College of Respiratory Therapists of Ontario, and the College of Nurses of Ontario; 1999.

24 Clarke M, Oxman AD, eds. *Cochrane Reviewers' Handbook 4.0* In: *Review Manager (RevMan)* [computer program]. Oxford, England: The Cochrane Library, The Cochrane Collaboration; 1999.

25 Van den Ende CHM, Vliet Vlietland TPM, Munneke M, Hazes JMW. *Dynamic Exercise Therapy for Rheumatoid Arthritis (Cochrane Review)* [Update software]. Oxford, England: The Cochrane Library, The Cochrane Collaboration; 2000:2.

26 Green S, Buchbinder R, Glazier RH, Forbes A. *Interventions for Shoulder Pain (Cochrane Review).* Oxford, England: The Cochrane Library, The Cochrane Collaboration; 1999:1.

27 Gadsby JG, Flowerdew MW. *The Effectiveness of Transcutaneous Electrical Nerve Stimulation (TENS) and Acupuncture-like Transcutaneous Electrical Nerve Stimulation (ALTENS) in the Treatment of Patients With Chronic Low Back Pain* [Update software]. Oxford, England: The Cochrane Library, The Cochrane Collaboration; 1997.

28 Gross AR, Aker PD, Goldsmith CH, et al. *Physical Medicine Modalities for Neck Pain* [Update software, updated quarterly on CD-ROM]. Oxford, England: The Cochrane Library, The Cochrane Collaboration; 1998:1.

29 Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ.* 1994;309(6964):1286–1291.

30 Haynes R, Wilczynski N, Mckibbon KA, Walker CJ. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *Journal of the American Medical Information Association.* 1994; 1:447–458.

31 Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized trials: is blinding necessary? *Control Clin Trials.* 1996;17:1–12.

32 Clark HD, Wells GA, Huet C, et al. Assessing the quality of randomized trials: reliability of the Jadad scale. *Control Clin Trials.* 1999;20: 448–452.

33 Canadian Task Force on the Periodic Health Examination. The periodic health examination: 1986 update. *Can Med Assoc J.* 1986; 134(721):729.

34 Browman GP, Newman TE, Mohide EA, et al. Progress of clinical oncology guidelines development using the Practice Guidelines Development Cycle: the role of practitioner feedback. *J Clin Oncol.* 1998;16: 1226–1231.

35 Guyatt GH, Sackett DL, Sinclair JC, et al. Users' guides to the medical literature, IX: method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA.* 1995;13;274: 1800–1804.

36 Felson DT, Anderson JJ, Boers M, et al. Preliminary definition of improvement in rheumatoid arthritis (American College of Rheumatology). *Arthritis Rheum.* 1995;38:727–735.

37 Stratford PW, Binkley JM, Solomon P, et al. Defining the minimum level of detectable change for the Roland-Morris Questionnaire. *Phys Ther.* 1996;76:359–368.

38 Rush PJ, Shore A. Physician perceptions of the value of physical modalities in the treatment of musculoskeletal disease. *Br J Rheumatol.* 1994;33:566–568.

39 Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet.* 1999;354:1896–1900.

40 Delitto A. Are measures of function and disability important in low back care? *Phys Ther.* 1994;74:452–462.

41 Deyo RA, Battié M, Beurskens AJ, et al. Outcome measures for low back pain research: a proposal for standardized use. *Spine.* 1998;23: 2003–2013.

42 Stratford PW, Binkley JM, Solomon P, et al. Assessing change over time in patients with low back pain. *Phys Ther.* 1994;74:528–533.

43 Stratford PW, Solomon P, Binkley JM, et al. Sensitivity of Sickness Impact Profile items to measure change over time in a low-back pain patient group. *Spine.* 1993;18:1723–1727.

44 Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol.* 1998;25: 198–199.

45 Deyo RA, Walsh NE, Schoenfeld LS, Ramamurthy S. Can trials of physical treatments be blinded? the example of transcutaneous electrical nerve stimulation for chronic pain. *Am J Phys Med Rehabil.* 1990;69:6–10.

46 Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther.* 1995;75:470–489.

47 Delitto A, Snyder-Mackler L. The diagnostic process: examples in orthopaedic physical therapy. *Phys Ther.* 1995;75:203–211.

48 Magistro C, Delitto A, Guccione AA, et al. Diagnosis in physical therapy. *PT Magazine.* 1993;1(6):58–65.

49 Zimny NJ, Goodman CC, Orest M, et al. Physical therapy diagnosis [letters to the editor]. *Phys Ther.* 1995;75:635–638.

50 Morin M, Brosseau L, Quirion-DeGrardi C. A theoretical framework on low level laser therapy (classes I, II, and III) application for the treatment of OA and RA. In: *Proceedings of the Canadian Physiotherapy Association National Congress.* 1996:1.

51 McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet.* 2000;356(9237):1228–1231.

52 Ottenbacher KJ. Why rehabilitation research does not work (as well as we think it should). *Arch Phys Med Rehabil.* 1995;76:123–129.

53 Gam AN, Johannsen F. Ultrasound therapy in musculoskeletal disorders: a meta-analysis. *Pain.* 1995;63:85–91.

54 Jette AM, Delitto A. Physical therapy treatment choices for musculoskeletal impairments. *Phys Ther.* 1997;77:145–154.

55 Duckworth M. Outcome measurement selection and typology. *Physiotherapy*. 1999;85:21–26.

56 DeRosa CP, Porterfield JA. A physical therapy model for the treatment of low back pain. *Phys Ther.* 1992;72:261–269.

57 Fritz H. Three-dimensional biomechanical model for simulating the response of the human body to vibration stress. *Med Biol Eng Comput.* 1998;36:686–692.

58 Schlapbach P, Gerber NJ. *Physiotherapy: Controlled Trials and Facts.* Basel, Switzerland: Karger; 1991

59 Graham ID, Calder LA, Hebert PC, et al. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care.* 2000;16:1024–1038.

60 Logan J, Graham ID. Toward a comprehensive interdisciplinary model of health care research use. *Science Communication.* 1998;20: 227–246.

Appendix.

Methodological Quality Checklist, Validated by Jadad et al[31]

Quality Scoring Information

	Yes	No
1. Was the study described as randomized (this includes the use of words such as "randomly," "random," and "randomization")?	☐	☐
2. Was the study described as double-blind?	☐	☐
3. Was there a description of withdrawals and dropouts?	☐	☐

Scoring the items:

Give a score of 1 for each "yes" and 0 points for each "no." There are no in-between marks.

Give 1 additional point if:

On question 1, the method of randomization was described and it was appropriate (eg, table of random numbers computer generated, coin tossing)
and/or
On question 2, the method of double-blinding was described and it was appropriate (eg, identical placebo, active placebo, dummy)

Deduct 1 point if:

On question 1, the method of randomization was described and it was inappropriate (eg, patients were allocated alternatively or according to date of birth, hospital number, etc)
and/or
On question 2, the study was described as double-blind but the method of blinding was inappropriate (eg, comparison of tablet versus injection with no double dummy)
Score: ————————