CHIROPRACTIC BEST PRACTICES
A Systematic Review by the Research Commission of the Council on Chiropractic Guidelines and Practice Parameters

METHODS

This document is solely a survey of existing studies, and only expresses the opinion of CCGPP. It is not intended to, nor does it establish a standard of care in specific communities, specific cases, or as to the care of any particular individual or condition. Each case must be determined on the basis of a careful clinical examination and diagnosis of the patient, giving due consideration to the specific condition presented and the individual's informed choice as to care and treatment. No part of this document is intended to support any litigation or proceeding involving the standard of care, medical necessity or reimbursement eligibility.

Intended audience:

- Chiropractors
- Chiropractic students and prospective students
- Chiropractic educators/educational institutions
- Chiropractic organizations/agencies
- Third-party payers
- Governmental agencies
- Patients and prospective patients

Overview of Process

The Council on Chiropractic Guidelines and Practice Parameters (CCGPP) and its organizational structure are described in the Introduction (http://www.ccgpp.org/introduction.pdf). Representing its constituent member organizations, the Council approved a listing of disorders by ICD9-CM codes that would form the scope of the investigations. CCGPP teams were identified, consisting of content experts from within the profession and involving consultants that are cross-trained or external to the profession in select areas.

The practice of chiropractic was divided into general areas based on anatomical regions (Table 1) and the list was given to the Commission who used it to design the literature searches within each domain. Using surveys of the profession[1,2] and publications on practice audits,[3-5] the team selected the topics for review by this first iteration. The criteria used were based on the team's determination of the most common disorders seen, and most common classifications of treatments used by chiropractors based on the literature.

Table 1: Practice domains identified for grouping of similar conditions for searching the literature and reporting of best practices.

Best Practice Domain	*Team Lead*
Introduction	John J. Triano, DC,PhD,FCCS(C)
Low back pain and related disorders of the lower extremities	William C. Meeker, DC, MPH
Neck pain, headache and neck related disorders of the upper extremities	Donald Murphy, DC
Thoracic spine and costovertebral joint disorders	Jeffrey Cates, DC, DABCCC
Upper extremity disorders	Thomas Souza, DC
Lower extremity disorders	Stephen Perle, DC, MS
Soft tissue disorders	Gordon Lawson, DC, MS
Wellness, non-musculoskeletal disorders, prevention and special populations	Cheryl Hawk, DC, PhD

Team selection and orientation training of team leaders

The CCGPP Council appointed two Co-chair for the Commission, each having experience in practice, structured literature review and formal consensus processes, either having been involved with one or more of original clinical and educational research, the Agency for Health Care Policy and Research acute low back pain guidelines, the RAND corporation task forces on appropriateness for use of spinal manipulation and earlier CCGPP Mercy Center chiropractic guidelines. Team leaders were nominated by the Commission co-chairmen and recommended to the Council for approval. Selection was based upon identification of individuals with clinical experience, additional cross-training in the content area of their assigned domain and / or scholarly work. Team members were selected from a multidisciplinary list of practitioners and content experts that had been solicited from the Council stakeholders and colleges. Additional nominees were identified to serve as consultants based on content expertise. Once a team leader accepted members for his/her team, no changes were

permitted in the team composition without being initiated by the team lead to add or replace members as necessary. All changes were submitted to the Council for agreement before implementation. All Commission member's service was uncompensated.

A team lead packet of information setting out motivation and methodology, including standardized instruments, with example formats for the final report was distributed. An orientation meeting was convened with all team leads and available consultants at the 2004 Association of Chiropractic Colleges Research Agenda Conference held in Las Vegas. Survey of the literature, rating and interpretation of evidence commenced in July of 2004 leading to this report.

Summary of Process

Balancing patient-centered and evidence-based values imparts similar internal tensions with tendency for the best intent of individuals to succumb to training biases and personal preferences. Four strategies were used to minimize this problem while empowering legitimate and informed interpretation of the literature:

1. Review of the literature by a panel of experts including those who do use and those who do not use the methods under review.
2. Standardized and validated, structured instruments for rating the quality of and results from the literature.
3. Formal consensus process, based on Delphi and Nominal Group Process, to adjudicate differences in professional opinion on the literature or to address important areas where literature is weak or lacking.
4. Wide stakeholder review with opportunity for critical comment offered to all stakeholder groups including patients, professionals, policymakers and third party payers.

A schematic of the process followed in developing the conclusions for best practices is given in Figure 1. Process development was guided by experience of Commission members with the RAND consensus process, Cochrane collaboration , AHCPR and published recommendations{National Health and Medical Research Council 1999 4118 /id} modified to the needs of the Council.

Methods used to Identify and Retrieve Evidence

- *Identification*
    - Topics were selected based on the most common disorders seen, and most common classifications of treatments used by chiropractors based on the literature.
- *Retrieval*
    - Searches of electronic databases
    - Hand-searches of published literature

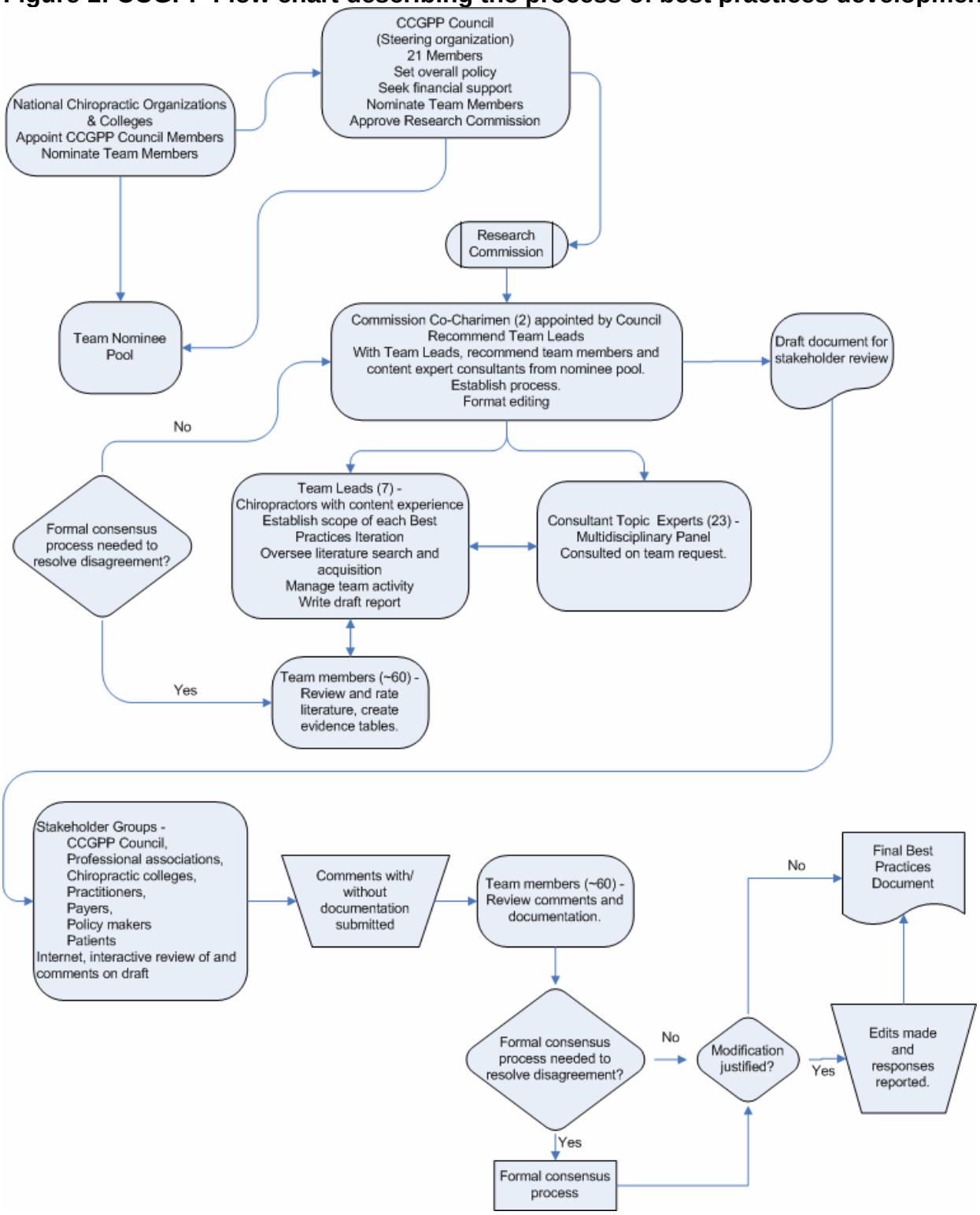
Methods to Evaluate the Evidence

- Standardized and validated instruments evaluating:
    - Meta-analyses & Systematic reviews
    - Randomized controlled trials (RCTs)
    - Cohort and case control studies
    - Case series (observational reports of more than 2 cases)
    - Diagnostic studies
- Standardized rating scheme
- Multidisciplinary panel conducting the review and rating

Topic Selection

Patients having many clinical descriptions seek care from chiropractors based upon the generally recognized reputation and the individual doctor's practice focus. Some providers center specifically on subluxation and its manifestations while others limit their practice to treating patients with spinal disorders or musculoskeletal complaints. Finally, others address more general health problems, prevention and special populations. The diversity of professional practice makes the review of all related literature to conclude evidence on Best Practices an impossible task. To accommodate the need for substantive review of the most relevant and informative literature, an interactive process was developed.

Figure 1: CCGPP Flow chart describing the process of best practices development.



Definitions for evidence ratings

Summary of grading of recommendations

GRADE A: Good evidence from relevant studies.
- Studies with appropriate designs and sufficient strength to answer the questions.
- Results are both clinically important and consistent with minor exceptions at most.
- Results are free of significant doubts about generalizability, bias, and design flaws.
- Negative studies have sufficiently large sample sizes to have adequate statistical power.

GRADE B: Fair evidence from relevant studies.
- Studies of appropriate designs of sufficient strength, but inconsistencies or minor doubts about generalizability, bias, and design flaws, or adequacy of sample size.
- Evidence solely from weaker designs, but confirmed in separate studies.

GRADE C: Limited evidence from studies/reviews.
- Studies with substantial uncertainty due to design flaws, or adequacy of sample size.
- Limited number of studies weak design for answering the question addressed.

GRADE D: Expert opinion, and usual and customary clinical practice.
Evidence consists of expert opinion; research cannot be or has not been performed.

GRADE I: No recommendation can be made because of insufficient or non-relevant evidence.
No evidence that directly pertains to the addressed question either because studies have not been performed or published, or are non-relevant.

GRADE A: Supported by good evidence from relevant studies. Must be included in evidence tables and as a reference(s) for best practices.

Explanation
- The evidence consists of results from studies based on appropriate research designs of sufficient strength to answer the questions addressed.
- Results are both clinically important and consistent with minor exceptions at most.
- The results are free of any significant doubts about generalizability, bias, and flaws in research design.
- Studies with negative results have sufficiently large sample sizes to have adequate statistical power.

Examples
- Supporting evidence may consist of a systematic review of randomized controlled trials (RCT's) with comparable methodology and consistent results or preponderance of evidence from several relevant RCT's with consistent results.
- For diagnostic tests - a systematic review of studies meeting standards of reporting diagnostic accuracy; or at least 1 study meeting standards of diagnostic accuracy, including cohort studies with good reference standards.
- For the question of natural history of a disorder, in the absence of evidence to the contrary, evidence might be a single well done prospective cohort study.

GRADE B: Supported by fair evidence from relevant studies. Must be included in evidence tables and as reference(s) for best practices.

Explanation

- The evidence consists of results from studies based on appropriate research designs of sufficient strength to answer the questions addressed, but there is some uncertainty attached to the conclusion because of inconsistencies among the results from the studies, or because of minor doubts about generalizability, bias, and research design flaws, or adequacy of sample size.
- Alternatively, the evidence consists solely of results from weaker designs for the question addressed, but the results have been confirmed in separate studies and are consistent with major exceptions at most.

Examples

- Supporting evidence might consist of a several RCT's with differing results although overall the results support the conclusion.
- The evidence might also be the result of a single randomized controlled trial with a clinically significant conclusion but doubtful generalizability.
- Alternatively, the evidence might come from a systematic review of RCT's with similar methodologies but differing results.
- For diagnostic tests, exploratory cohort studies with good reference standards, or instrumentation studies of reliability and validity.
- For a question of harm or adverse events, the evidence might consist of 2 or more independent case control studies with similar conclusions and minimal bias and research design flaws.

GRADE C: Supported by limited evidence from studies or reviews. Do not include in evidence tables but as reference(s) for best practices.

Explanation

- The evidence consists of results from studies of appropriate design for answering the question addressed, but there is substantial uncertainty attached to the conclusions because of inconsistencies among the results from different studies, or because of serious doubts about generalizability, bias, research design flaws, or adequacy of sample size.
- Alternatively, the evidence consists solely of results from a limited number of studies or because of weak design for answering the question addressed.

Examples

- For a question of treatment efficacy or effectiveness, the evidence might consist of systematic or narrative reviews or RCT's with contradictory results and/or serious methodological flaws.
- From relevant cohort, case control, ecological studies, and outcomes research.
- Alternately, the evidence might consist of individual case series.
- For diagnostic studies, the evidence might consist of non-consecutive studies without appropriate reference standards and case control studies unconfirmed by other studies.
- For a question or harm, the evidence might consist of results from a single case control study, or case series.

GRADE D:
Details: Supported by expert opinion, and usual and customary clinical practice. Include as reference(s) for best practices.
Explanation

- The evidence consists of expert opinion. Research studies cannot be or have not been performed.

Examples

- The literature cited might consist of a consensus report, a consensus opinion based on practice guidelines, an editorial, a position statement from a national body without citations of the results of research studies, and single case reports.

GRADE I: No recommendation can be made because of insufficient or non-relevant evidence. It should not be included in evidence tables or as reference(s) for best practices.
Explanation

- There is no evidence that directly pertains to the addressed question because either the studies have not been performed or published, or are non-relevant.

Examples

- No studies could be identified using optimal search strategies of appropriate data bases, or by hand searching. Alternately, the literature cited does not have direct bearing on the question being addressed.

Use of Evidence Tables

Evidence tables for RCTs rated by the team were constructed using categorical information shown reliable in other studies. Templates were provided to each team member for recording this information during the course of their review.

Stakeholder review and implementation

Stakeholder review of best practices is a critical step to facilitate final recommendations and implementation. This process affords the opportunity for individuals and groups that can be impacted by best practices to provide comment and documentation for consideration by the team. Stakeholders for the low back and related lower extremity symptoms are considered to include Doctors of Chiropractic, students and prospective students, educators and teaching institutions, professional organizations and agencies, third-party payers, governmental agencies and patients.

Three separate strategies have been used to inform interprofessional stakeholders on progress during the development of the best practices document. By providing periodic updates, colleges, associations and providers were made aware of the pending release for review and comment. The three methods included 1) periodic articles published in interprofessional news media, 2) presentations at the Association of Chiropractic Colleges, the Federation of Chiropractic Licensing Boards meetings, and 3) providing a speaker's bureau for use in presentations to state professional association meetings.

Two strategies were used to reach stakeholders for review and comment on the document itself. On completion of the draft document of best practices, a summary of the best practices document was posted on a widely accessed health care web site (Spine-health.com) that experienced a public hit rate of 2.5 to 3.0 million per month during 2005. Separately, on the CCGPP web site, the document was posted and notification made to colleges, state and national associations and third-party payers.

Interactive electronic questionnaires, developed by the Dissemination, Implementation, Evaluation and Review (DIER) committee of CCGPP are available for stakeholder comments on-line. Those choosing to comment are invited to submit documentation for their opinions directly to CCGPP. The postings will be maintained for 60 days and comments harvested electronically and provided to the co-chair of the Commission. The co-chair will group similar comments and develop summary questions that will be posed, with the original comments and any supportive documentation, to the team for review and response. A tally of comments by group along with the questions and responses from the team will be made a part of the Appendix in the final document release.

The final document will reflect any changes in conclusions of the team made in response to stakeholder input.

Audit and Review

As noted earlier, the best practices effort of the CCGPP is designed as an iterative process. The low back and related lower extremity best practices document is intended to be reviewed with inclusion of any new evidence and extension of the domains considered on a 2 to 5 year cycle, depending on the state of the art in the literature. (Details on Audit and Review to be completed in conference with the DIER committee following the stakeholder review and comments.)

References

1. McDonald WP DK, Pfefer M. How Chiropractors Think and Practice: The Survey of North American Chiropractors. *Seminars in Integrative Medicine.* September 2004 2004;2(3):92-98.
2. Christensen M, Kollasch M, Ward R, Webb K, Day A, ZumBrunnen J. *Job Analysis of Chiropractic*. Greeley, CO: NBCE; 2005.
3. Coulter ID, Hurwitz EL, Adams AH, Genovese BJ, Hays R, Shekelle PG. Patients using chiropractors in North America: who are they, and why are they in chiropractic care? *Spine.* Feb 1 2002;27(3):291-296; discussion 297-298.
4. Coulter ID, Singh BB, Riley D, Der-Martirosian C. Interprofessional referral patterns in an integrated medical system. *J Manipulative Physiol Ther.* Mar-Apr 2005;28(3):170-174.
5. Hurwitz EL, Coulter ID, Adams AH, Genovese BJ, Shekelle PG. Use of chiropractic services from 1985 through 1991 in the United States and Canada. *Am J Public Health.* May 1998;88(5):771-776.

Evaluation Checklists Used in Literature Synthesis Process

SIGN	Methodology Checklist 1: Systematic Reviews and Meta-analyses

Study identification (*Include author, title, year of publication, journal title, pages*)

Guideline topic:	Key Question No:

Checklist completed by:

Section 1: Internal validity

In a well conducted systematic review		**In this study this criterion is::**	
1.1	The study addresses an appropriate and clearly focused question.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.2	A description of the methodology used is included.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	***The literature search is sufficiently rigorous to identify all the relevant studies.***	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	Study quality is assessed and taken into account.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.5	There are enough similarities between the studies selected to make combining them reasonable.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

SECTION 2: OVERALL ASSESSMENT OF THE STUDY

	How well was the study done to minimise bias? *How valid is the study?* Code +, n, or –	

Notes on the use of Methodology Checklist 1: Systematic Reviews and Meta-analyses

Section 1 identifies the study, the reviewer, the guideline for which the paper is being considered as evidence, and the key question(s) it is expected to address. The reviewer is asked to consider a series of aspects of study design and to make a judgment as to how well the current study meets each criterion. Each relates to an aspect of methodology that research has shown to be likely to influence the conclusions of a study.

For each question in this section you should use one of the following to indicate how well it has been addressed in the study:

- Well covered
- Adequately addressed
- Poorly addressed
- Not addressed *(i.e. not mentioned, or indicates that this aspect of study design was ignored)*
- Not reported *(i.e. mentioned, but insufficient detail to allow assessment to be made)*
- Not applicable.

1.1 *The study addresses an appropriate and clearly focused question.*

Unless a clear and well defined question is specified in the report of the review, it will be difficult to assess how well it has met its objectives or how relevant it is to the question you are trying to answer on the basis of the conclusions.

1.2 *A description of the methodology used is included.*
One of the key distinctions between a systematic review and a general review is the systematic methodology used. A systematic review should include a detailed description of the methods used to identify and evaluate individual studies. If this description is not present, it is not possible to make a thorough evaluation of the quality of the review, and **it should be rejected as a source of Level 1 evidence**. (Though it may be useable as Level 4 evidence, if no better evidence can be found.)

1.3 *The literature search is sufficiently rigorous to identify all the relevant studies.*

A systematic review based on a limited literature search – e.g. one limited to Medline only – is likely to be heavily biased. A well conducted review should as a minimum look at Embase and Medline, and from the late 1990s onward, the Cochrane Library. Any indication that hand searching of key journals, or follow up of reference lists of included studies were carried out in addition to electronic database searches can be taken as evidence of a well conducted review.

1.4 *Study quality is assessed and taken into account.*

A well conducted systematic review should have used clear criteria to assess whether individual studies had been well conducted before deciding whether to include or exclude them. If there is no indication of such an assessment, the individual papers included in the review must be obtained and their methodology evaluated.

1.5 *There are enough similarities between the studies selected to make combining them reasonable.*
Studies covered by a systematic review should be selected using clear inclusion criteria. These criteria should include, either implicitly or explicitly, the question of whether the selected studies can legitimately be compared. It should be clearly ascertained, for example, that the populations covered by the studies are comparable; that the methods used in the investigations are the same; that the outcome measures are comparable; and the variability in effect sizes between studies is not greater than would be expected by chance alone.

Section 2 relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

+	**Strong. All or most** of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
n	If the answers to the questions pertaining to PLUS or MINUS criteria do not indicate that the report is exceptionally strong or exceptionally weak, the report should be designated with an n for neutral
-	**Weak. Few or no** criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

SIGN	Methodology Checklist 2: Randomised Controlled Trials

Study identification (*Include author, title, year of publication, journal title, pages*)

Guideline topic:	Key Question No:

Checklist completed by:

Section 1: Internal validity

In a well conducted RCT study…..		**In this study this criterion is::**
1.1	The study addresses an appropriate and clearly focused question.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.2	The assignment of subjects to treatment groups is randomised	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.3	***An adequate concealment method is used***	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.4	Subjects and investigators are kept 'blind' about treatment allocation	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.5	The treatment and control groups are similar at the start of the trial	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.6	The only difference between groups is the treatment under investigation	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.7	All relevant outcomes are measured in a standard, valid and reliable way	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.8	What % of individuals recruited into each treatment arm of the study dropped out before the study was completed?	
1.9	*All the subjects analysed in groups to which they were randomly allocated (e.g. intention to treat analysis)*	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.10	Where the study is carried out at more than one site, results are comparable for all sites	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable

SECTION 2: OVERALL ASSESSMENT OF THE STUDY

	How well was the study done to minimise bias? How valid is the study? Code +, n, or –	

Notes on the use of Methodology Checklist 2: Randomised Controlled Trials

Section 1 identifies the study, the reviewer, the guideline for which the paper is being considered as evidence, and the key question(s) it is expected to address. The reviewer is asked to consider a series of aspects of RCT design and to make a judgement as to how well the current study meets this criterion. Each relates to an aspect of methodology that research has shown makes a significant difference to the conclusions of a study.

For each question in this section you should use one of the following to indicate how well it has been addressed in the study:

- Well covered
- Adequately addressed
- Poorly addressed
- Not addressed *(i.e. not mentioned, or indicates that this aspect of study design was ignored)*
- Not reported *(i.e. mentioned, but insufficient detail to allow assessment to be made)*
- Not applicable.

1.1 *The study addresses an appropriate and clearly focused question*

Unless a clear and well defined question is specified, it will be difficult to assess how well the study has met its objectives or how relevant it is to the question you are trying to answer on the basis of its conclusions.

1.2 *The assignment of subjects to treatment groups randomised*

Random allocation of patients to receive one or other of the treatments under investigation, or to receive either treatment or placebo, is fundamental to this type of study. **If there is no indication of randomisation, the study should be rejected.** If the description of randomisation is poor, the study should be given a lower quality rating. Processes such as alternate allocation, allocation by date of birth, or day of the week attending a clinic are not true randomisation processes and it is easy for a researcher to work out which patients received which treatment. These studies should therefore be classed as Controlled Clinical Trials rather than RCTs.

1.3 *An adequate concealment method is used*

Allocation concealment refers to the process used to ensure that researchers are unaware which group patients are being allocated to at the time they enter the study. Research has shown that where allocation concealment is inadequate, investigators can overestimate the effect of interventions by up to 40%. Centralised allocation, computerised allocation systems, or the use of coded identical containers would all be regarded as adequate methods of concealment, and may be taken as indicators of a well conducted study. If the method of concealment used is regarded as poor, or relatively easy to subvert, the study must be given a lower quality rating, and can be rejected if the concealment method is seen as inadequate.

1.4 *Subjects and investigators are kept 'blind' to treatment allocation*
Blinding refers to the process whereby people are kept unaware of which treatment an individual patient has been receiving when they are assessing the outcome for that patient. It can be carried out up to three levels. Single blinding is where patients are unaware of which treatment they are receiving. In double blind studies neither the doctor nor the patient knows which treatment is being given. In very rare cases studies may be triple blinded, where neither patients, doctors, nor those conducting the analysis are aware of which patients received which treatment. The higher the level of blinding, the lower the risk of bias in the study.

1.5 *The treatment and control groups were similar at the start of the trial*

Patients selected for inclusion in a trial must be as similar as possible. The study should report any significant differences in the composition of the study groups in relation to gender mix, age, stage of disease (if appropriate), social background, ethnic origin, or comorbid conditions. These factors may be covered by inclusion and exclusion criteria, rather than being reported directly. Failure to address this question, or the use of inappropriate groups, should lead to the study being downgraded.

1.6 *The only difference between the groups is the treatment under investigation*

If some patients received additional treatment, even if of a minor nature or consisting of advice and counselling rather than a physical intervention, this treatment is a potential confounding factor that may invalidate the results. **If groups were *not* treated equally, the study should be rejected unless no other evidence is available.** If the study is used as evidence it should be treated with caution.

1.7 *All relevant outcomes measured in a standard, valid and reliable way*
The primary outcome measures used should be clearly stated in the study. **If the outcome measures are not stated, or the study bases its main conclusions on secondary outcomes, the study should be rejected.** Where outcome measures require any degree of subjectivity, some evidence should be provided that the measures used are reliable and have been validated prior to their use in the study.
1.8 *What percentage of the individuals or clusters recruited into each treatment arm of the study dropped out before the study was completed?*
The number of patients that drop out of a study should give concern if the number is very high. Conventionally, a 20% drop out rate is regarded as acceptable, but this may vary. Some regard should be paid to why patients dropped out, as well as how many. It should be noted that the drop out rate may be expected to be higher in studies conducted over a long period of time. A higher drop out rate will normally lead to downgrading, rather than rejection of a study.
1.9 *All the subjects are analysed in the groups to which they were randomly allocate (intention to treat analysis)*
In practice, it is rarely the case that all patients allocated to the intervention group receive the intervention throughout the trial, or that all those in the comparison group do not. Patients may refuse treatment, or contra-indications arise that lead them to be switched to the other group. If the comparability of groups through randomisation is to be maintained, however, patient outcomes must be analysed according to the group to which they were originally allocated irrespective of the treatment they actually received. (This is known as intention to treat analysis.) If it is clear that analysis was not on an intention to treat basis, the study may be rejected. If there is little other evidence available, the study may be included but should be evaluated as if it were a non-randomised cohort study.

1.10 *Where the study is carried out at more than one site, results are comparable for all sites*

In multi-site studies, confidence in the results should be increased if it can be shown that similar results were obtained at the different participating centres.

Section 2 relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

+	**Strong. All or most** of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
n	If the answers to the questions pertaining to PLUS or MINUS criteria do not indicate that the report is exceptionally strong or exceptionally weak, the report should be designated with an n for neutral
-	**Weak. Few or no** criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

SIGN	Methodology Checklist 3: Cohort studies

Study identification (Include author, title, year of publication, journal title, pages)	
Guideline topic:	Key Question No:
Checklist completed by:	

Section 1: Internal validity

In a well conducted cohort study:		**In this study the criterion is:**
1.1	The study addresses an appropriate and clearly focused question.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
SELECTION OF SUBJECTS		
1.2	The 2 groups are selected from source populations comparable in all respects other than the factor under investigation.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.3	The study indicates how many of the people asked to take part did so, in each of the groups being studied.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.4	The likelihood that some eligible subjects might have the outcome at time of enrollment is assessed/accounted for in analysis.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.5	What percentage of individuals or clusters recruited into each arm of the study dropped out before the study was completed.	
1.6	***Comparison is made between full participants and those lost to follow up, by exposure status.***	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
ASSESSMENT		
1.7	The outcomes are clearly defined.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.8	The assessment of outcome is made blind to exposure status.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.9	Where blinding was not possible, there is recognition that knowledge of exposure could have influenced outcome assessment.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.10	The measure of assessment of exposure is reliable.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable
1.11	Evidence from other sources is used to demonstrate that the method of outcome assessment is valid and reliable.	Well covered / adequately addressed / poorly addressed / not addressed / not reported / not applicable

1.12	Exposure level or prognostic factor is assessed more than once.	Well covered adequately addressed poorly addressed not addressed not reported not applicable
CONFOUNDING		
1.13	The main potential confounders are identified and taken into account in the design and analysis.	Well covered adequately addressed poorly addressed not addressed not reported not applicable
STATISTICAL ANALYSIS		
1.14	Have confidence intervals been provided?	
SECTION 2: OVERALL ASSESSMENT OF THE STUDY		
	How well was the study done to minimise bias? How valid is the study? Code +, n, or –	

Notes on the use of Methodology Checklist 3: Cohort studies

The studies covered by this checklist are designed to answer questions of the type "What are the effects of this exposure?", It relates to studies that compare a group of people with a particular exposure with another group who either have not had the exposure, or have a different level of exposure. Cohort studies may be prospective (where the exposure is defined and subjects selected before outcomes occur), or retrospective (where exposure is assessed after the outcome is known, usually by the examination of medical records). Retrospective studies are generally regarded as a weaker design, and should not receive a "++" rating.

Section 1 identifies the study, the reviewer, the guideline for which the paper is being considered as evidence, and the key question(s) it is expected to address. The reviewer is asked to consider a series of aspects of cohort study design and to make a judgement as to how well the current study meets this criterion. Each relates to an aspect of methodology that research has shown to be likely to influence the conclusions of a study.

Because of the potential complexity and subtleties of the design of this type of study, there are comparatively few criteria that automatically rule out use of a study as evidence. It is more a matter of increasing confidence in the strength of association between exposure and outcome by identifying how many aspects of good study design are present, and how well they have been tackled. A study that fails to address or report on more than one or two of the questions addressed below should almost certainly be rejected.

For each question in this section you should use one of the following to indicate how well it has been addressed in the study:

- Well covered
- Adequately addressed
- Poorly addressed
- Not addressed (i.e. not mentioned, or indicates that this aspect of study design was ignored)
- Not reported (i.e. mentioned, but insufficient detail to allow assessment to be made)
- Not applicable.

1.1 *The study addresses an appropriate and clearly focused question?*
Unless a clear and well defined question is specified, it will be difficult to assess how well the study has met its objectives or how relevant it is to the question you are trying to answer on the basis of its conclusions.

1.2 *The two groups being studied are selected from source populations that are comparable in all respects other than the factor under investigation.*

It is important that the two groups selected for comparison are as similar as possible in all characteristics except for their exposure status, or the presence of specific prognostic factors or prognostic markers relevant to the study in question. **If the study does not include clear definitions of the source populations and eligibility criteria for participants it should be rejected.**

1.3 *The study indicates how many of the people asked to take part did so, in each of the groups being studied.*
The participation rate is defined as the number of study participants divided by the number of eligible subjects, and should be calculated separately for each branch of the study. A large difference in participation rate between the two arms of the study indicates that a significant degree of selection bias may be present, and the study results should be treated with considerable caution.

1.4 *The likelihood that some eligible subjects might have the outcome at the time of enrolment is assessed and taken into account in the analysis?*

If some of the eligible subjects, particularly those in the unexposed group, already have the outcome at the start of the trial the final result will be biased. A well conducted study will attempt to estimate the likelihood of this occurring, and take it into account in the analysis through the use of sensitivity studies or other methods.

1.5 *What percentage of individuals or clusters recruited into each arm of the study dropped out before the study was completed?*

The number of patients that drop out of a study should give concern if the number is very high. Conventionally, a 20% drop out rate is regarded as acceptable, but in observational studies conducted over a lengthy period of time a higher drop out rate is to be expected. A decision on whether to downgrade or reject a study because of a high drop out rate is a matter of judgement based on the reasons why people dropped out, and whether drop out rates were comparable in the exposed and unexposed groups. Reporting of efforts to follow up participants that dropped out may be regarded as an indicator of a well conducted study.

1.6 *Comparison is made between full participants and those lost to follow-up, by exposure status.*

For valid study results, it is essential that the study participants are truly representative of the source population. It is always possible that participants who dropped out of the study will differ in some significant way from those who remained part of the study throughout. A well conducted study will attempt to identify any such differences between full and partial participants in both the exposed and unexposed groups. Any indication that differences exist, should lead to the study results being treated with caution.

1.7 *The outcomes are clearly defined.*

Once enrolled in the study, participants should be followed until specified end points or outcomes are reached. In a study of the effect of exercise on the death rates from heart disease in middle aged men, for example, participants might be followed up until death, or until reaching a predefined age. **If outcomes and the criteria used for measuring them are not clearly defined, the study should be rejected.**

1.8 *The assessment of outcome is made blind to exposure status*

If the assessor is blinded to which participants received the exposure, and which did not, the prospects of unbiased results are significantly increased. Studies in which this is done should be rated more highly than those where it is not done, or not done adequately.

1.9 *Where blinding was not possible, there is some recognition that knowledge of exposure status could have influenced the assessment of outcome.*

Blinding is not possible in many cohort studies. In order to asses the extent of any bias that may be present, it may be helpful to compare process measures used on the participant groups - e.g. frequency of observations, who carried out the observations, the degree of detail and completeness of observations. If these process measures are comparable between the groups, the results may be regarded with more confidence.

1.10 *The measure of assessment of exposure is reliable.*

A well conducted study should indicate how the degree of exposure or presence of prognostic factors or markers was assessed. Whatever measures are used must be sufficient to establish clearly that participants have or have not received the

exposure under investigation and the extent of such exposure, or that they do or do not possess a particular prognostic marker or factor. Clearly described, reliable measures should increase the confidence in the quality of the study.

1.11 *Evidence from other sources is used to demonstrate that the method of outcome assessment is valid and reliable.*

The primary outcome measures used should be clearly stated in the study. **If the outcome measures are not stated, or the study bases its main conclusions on secondary outcomes, the study should be rejected.** Where outcome measures require any degree of subjectivity, some evidence should be provided that the measures used are reliable and have been validated prior to their use in the study.

1.12 *Exposure level or prognostic factor is assessed more than once.*

Confidence in data quality should be increased if exposure level is measured more than once in the course of the study. Independent assessment by more than one investigator is preferable.

1.13 *The main potential confounders are identified and taken into account adequately in the design and analysis.*

Confounding is the distortion of a link between exposure and outcome by another factor that is associated with both exposure and outcome. The possible presence of confounding factors is one of the principal reasons why observational studies are not more highly rated as a source of evidence. The report of the study should indicate which potential confounders have been considered, and how they have been assessed or allowed for in the analysis. Clinical judgement should be applied to consider whether all likely confounders have been considered. If the measures used to address confounding are considered inadequate, the study should be downgraded or rejected, depending on how serious the risk of confounding is considered to be. **A study that does not address the possibility of confounding should be rejected.**

1.14 *Confidence intervals are provided.*

Confidence limits are the preferred method for indicating the precision of statistical results, and can be used to differentiate between an inconclusive study and a study that shows no effect. Studies that report a single value with no assessment of precision should be treated with extreme caution.

Section 2 relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

+	**Strong. All or most** of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
n	If the answers to the questions pertaining to PLUS or MINUS criteria do not indicate that the report is exceptionally strong or exceptionally weak, the report should be designated with an n for neutral
-	**Weak. Few or no** criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

.

SIGN	**Methodology Checklist 5: Studies of Diagnostic Accuracy**

Study identification (*Include author, title, reference, year of publication*)

Guideline topic:	Key Question No:

Checklist completed by:

SECTION 1: INTERNAL VALIDITY

	In a well conducted diagnostic study…	**In this study this criterion is**	
1.1	The nature of the test being studied is clearly specified.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.2	The test is compared with an appropriate gold standard.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	**Where no gold standard exists, a validated reference standard is used as comparator.**	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	**Patients for testing are selected either as a consecutive series or randomly, from a clearly defined study population.**	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.5	**The test and gold standard are measured independently (blind) of each other.**	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.6	The test and gold standard are applied as close together in time as possible.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.7	Results are reported for all patients that are entered into the study.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.8	**A pre-test diagnosis is made and reported.**	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

SECTION 2: OVERALL ASSESSMENT OF THE STUDY

	How well was the study done to minimise bias? How valid is the study? *Code +, n, or –*	

Notes on the use of Methodology Checklist 5: Diagnostic studies

This checklist is designed for the evaluation of studies assessing the accuracy of specific diagnostic tests. It does ***not*** address questions of the usefulness of the test in practice, or how the test compares with alternatives. These and other questions addressing the relevance of the test are entirely appropriate for guideline developers to consider, but form part of the considered judgment process where developers consider their interpretation of the evidence.

Section 1 Asks a series of questions aimed at establishing the internal validity of the study under review - i.e. making sure that it has been carried out carefully, and that the conclusions represent an unbiased assessment of the accuracy and reliability of the test being evaluated. Each question covers an aspect of methodology that has been shown to make a significant difference to the reliability of a study.

For each question in this section you should use one of the following to indicate how well it has been addressed in the study:

- Well covered
- Adequately addressed
- Poorly addressed
- Not addressed (i.e. not mentioned, or indicates that this aspect of study design was ignored)
- Not reported (i.e. mentioned, but insufficient detail to allow assessment to be made)
- Not applicable.

1.1 *The nature of the test being studied is clearly specified.*

The clinical protocol used in deciding when to use the test should be described. The rules by which observations made during the test are converted to a positive or negative result should be stated (e.g if a continuous variable is measured the threshold value should be given).

1.2 *The test is compared with an appropriate gold standard.*

In order to assess how well a new or alternative diagnostic test performs, it has to be compared with a reference standard so that the investigator has a clear idea of how effective it is at identifying cases or non-cases among the target population. The reference standard will be some form of existing test or diagnostic method whose accuracy is known within clearly defined limits. There should be an indication in the study of exactly what test was evaluated, what standard was used for comparison, and what evidence there is that the comparator is a valid one for the test under investigation.

The comparator should ideally be a "gold standard" that is accepted as giving a correct diagnosis (necessary, but not sufficient, evidence for this might include demonstrating that inter-observer variability can be assumed to be close to zero). Where a comparator other than a gold standard is used, it must be well characterised in terms of sensitivity and specificity.

In the checklist, and throughout these notes, "test" refers to the diagnostic test being evaluated. "Gold standard" refers to the standard against which the new test is being compared. **Where a gold standard exists any evaluation of a new test that does not make comparison with that standard should be rejected as evidence, unless a clear and justifiable explanation is provided as to why the gold standard was not used.** Only studies that have used a gold standard as comparator can be considered as high quality (++) evidence.

1.3 *Where no gold standard exists, a validated reference standard is used as comparator.*

If there is no applicable gold standard, the new test must be compared with an existing test of a known sensitivity and specificity. The study should justify the use of the selected comparator. The uncertainty in the diagnosis made with a non-gold standard should be taken account of in the analysis – it is not sufficient to calculate sensitivity and specificity under the assumption that the standard has provided a true diagnosis. Studies that do not use a gold standard as comparator cannot be rated as higher than moderate (+) evidence.

1.4 *Patients for testing are selected either consecutively or randomly, from a clearly defined study population.*

The critical point here is to ensure that selection of patients is not influenced by the likelihood of obtaining a particular result from the test. There should be clearly stated criteria defining which patients are part of the study population, and which are excluded.

1.5 *The test and gold standard are measured independently (blind) of each other.*

The test under investigation and the gold standard should both be applied to the same patients to allow results to be compared. Investigators carrying out each test should be blind to the results obtained from the other to ensure their evaluation of test results is not biased. The test being evaluated should normally be carried out before the gold standard. In some studies, not all patients receiving the test will also receive the gold standard. Measures must be taken to ensure that the choice of patients for testing with the gold standard is not influenced by the results of earlier tests. **Where such measures have not been taken or are deemed to be inadequate the study should be downgraded or rejected.**

1.6 *The test and gold standard are applied as close together in time as possible*

If too much time is allowed to pass between the application of the test and gold standard, the patient's condition is likely to change (particularly if an intervention has been introduced following the test result). The test and gold standard should therefore both be applied on the same day if possible, and if not the gold standard should be applied as soon as possible thereafter. Where a length of time has been allowed to pass between tests, the study should be downgraded (or rejected if the time lapse is too long to be justifiable).

1.7 *Results are reported for all patients that are entered into the study.*

Outcomes for all patients entered into the study should be reported, including any for whom test results are unavailable for any reason. Where a significant number (>20% as a guide) of patients do not have any reported results, the study should be downgraded or rejected.

1.8 *A pre-test diagnosis is made and reported.*

In order to evaluate the additional information available from a test it is necessary to know what diagnosis would have been made (and consequently what clinical action taken) in the absence of the test. This requires a pre-test diagnosis to be recorded in each trial subject.

Section 2 relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

+	**Strong. All or most** of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
n	If the answers to the questions pertaining to PLUS or MINUS criteria do not indicate that the report is exceptionally strong or exceptionally weak, the report should be designated with an n for neutral
-	**Weak. Few or no** criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

SIGN	Methodology Checklist 4: Case-control studies

Study identification (*Include author, title, year of publication, journal title, pages*)

Guideline topic:	Key Question No:

Checklist completed by:

Section 1: Internal validity

In an well conducted case control study:		**In this study the criterion is:**	
1.1	The study addresses an appropriate and clearly focused question	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
SELECTION OF SUBJECTS			
1.2	The cases and controls are taken from comparable populations	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	The same exclusion criteria are used for both cases and controls	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	What percentage of each group (cases and controls) participated in the study?	Cases: Controls:	
1.5	Comparison is made between participants and non-participants to establish their similarities or differences	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.6	Cases are clearly defined and differentiated from controls	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.7	**It is clearly established that controls are non-cases**	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
ASSESSMENT			
1.8	Measures will have been taken to prevent knowledge of primary exposure influencing case ascertainment	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

1.9	Exposure status is measured in a standard, valid and reliable way	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
CONFOUNDING			
1.10	The main potential confounders are identified and taken into account in the design and analysis	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
STATISTICAL ANALYSIS			
1.11	Confidence intervals are provided		
SECTION 2: OVERALL ASSESSMENT OF THE STUDY			
	How well was the study done to minimise bias? How valid is the study? Code +, n, or –		

Notes on the use of Methodology Checklist 4: Case-control studies

The studies covered by this checklist are designed to answer questions of the type "What are the factors that caused this event?", and involve comparison of individuals with an outcome with other individuals from the same population who do not have the outcome. These studies start after the outcome of an event, and can be used to assess multiple causes of a single event. They are generally used to assess the causes of a new problem, but may also be useful for the evaluation of population based interventions such as screening.

Section 1 identifies the study, the reviewer, the guideline for which the paper is being considered as evidence, and the key question(s) it is expected to address. The reviewer is asked to consider a series of aspects of cohort study design and to make a judgement as to how well the current study meets this criterion. Each relates to an aspect of methodology that research has shown makes a significant difference to the conclusions of a study.

Case-control studies need to be very carefully designed, and the complexity of their design is often not appreciated by investigators, leading to many poor quality studies being conducted. The questions in this checklist are designed to identify the main features that should be present in a well designed study. There are few criteria that should, alone and unsupported, lead to rejection of a study. However, a study that fails to address or report on more than one or two of the questions addressed below should almost certainly be rejected.

For each question in this section you should use one of the following to indicate how well it has been addressed in the study:

- Well covered
- Adequately addressed
- Poorly addressed
- Not addressed (i.e. not mentioned, or indicates that this aspect of study design was ignored)
- Not reported (i.e. mentioned, but insufficient detail to allow assessment to be made)
- Not applicable.

1.1 *The study addresses an appropriate and clearly focused question*

Unless a clear and well defined question is specified, it will be difficult to assess how well the study has met its objectives or how relevant it is to the question you are trying to answer on the basis of its conclusions.

1.2 *The cases and controls are taken from comparable populations.*

Study participants may be selected from the target population (all individuals to which the results of the study could be applied), the source population (a defined subset of the target population from which participants are selected), or from a pool of eligible subjects (a clearly defined and counted group selected from the source population. **If the study does not include clear definitions of the source population it should be rejected.**

1.3 *The same exclusion criteria are used for both cases and controls*

All selection and exclusion criteria should be applied equally to cases and controls. Failure to do so may introduce a significant degree of bias into the results of the study.

1.4 *What percentage of each group (cases and controls) participated in the study?*

Differences between the eligible population and the participants are important, as they may influence the validity of the study. A participation rate can be calculated by dividing the number of study participants by the number of eligible subjects. It is more useful if calculated separately for cases and controls. If the participation rate is low, or there is a large difference between the two groups, the study results may well be invalid due to differences between participants and non-participants. In these circumstances, the study should be downgraded, and rejected if the differences are very large.

1.5 *Comparison is made between participants and non-participants to establish their similarities or differences*

Even if participation rates are comparable and acceptable, it is still possible that the participants selected to act as cases or controls may differ from other members of the source population in some significant way. A well conducted case-control study will look at samples of the non-participants among the source population to ensure that the participants are a truly representative sample.

1.6 *Cases are clearly defined and differentiated from controls*

The method of selection of cases is of critical importance to the validity of the study. Investigators have to be certain that cases are truly cases, but must balance this with the need to ensure that the cases admitted into the study are representative of the eligible population. **The issues involved in case selection are complex, and should ideally be evaluated by someone with a good understanding of the design of case-control studies. If the study does not comment on how cases were selected, it is probably safest to reject it as a source of evidence.**

1.7 *It is clearly established that controls are non-cases*

Just as it is important to be sure that cases are true cases, it is important to be sure that controls do not have the outcome under investigation. Control subjects should be chosen so that information on exposure status can be obtained or assessed in a similar way to that used for the selection of cases. If the methods of control selection are not described, the study should be rejected. **If different methods of selection are used for cases and controls the study should be evaluated by someone with a good understanding of the design of case-control studies.**

1.8 *Measures will have been taken to prevent knowledge of primary exposure influencing case ascertainment*

If there is a possibility that case ascertainment can be influenced by knowledge of exposure status, assessment of any association is likely to be biased. A well conducted study should take this into account in the design of the study.

1.9 *Exposure status is measured in a standard, valid and reliable way*

The primary outcome measures used should be clearly stated in the study. **If the outcome measures are not stated, or the study bases its main conclusions on secondary outcomes, the study should be rejected.** Where outcome measures require

any degree of subjectivity, some evidence should be provided that the measures used are reliable and have been validated prior to their use in the study.

1.10 *The main potential confounders are identified and taken into account in the design and analysis*

Confounding is the distortion of a link between exposure and outcome by another factor that is associated with both exposure and outcome. The possible presence of confounding factors is one of the principal reasons why observational studies are not more highly rated as a source of evidence. The report of the study should indicate which potential confounders have been considered, and how they have been assessed or allowed for in the analysis. Clinical judgement should be applied to consider whether all likely confounders have been considered. If the measures used to address confounding are considered inadequate, the study should be downgraded or rejected, depending on how serious the risk of confounding is considered to be. **A study that does not address the possibility of confounding should be rejected.**

1.11 *Confidence intervals are provided*

Confidence limits are the preferred method for indicating the precision of statistical results, and can be used to differentiate between an inconclusive study and a study that shows no effect. Studies that report a single value with no assessment of precision should be treated with extreme caution.

Section 2 relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

+	**Strong. All or most** of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
n	If the answers to the questions pertaining to PLUS or MINUS criteria do not indicate that the report is exceptionally strong or exceptionally weak, the report should be designated with an n for neutral
-	**Weak. Few or no** criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

SIGN	Methodology Checklist 6: Economic Evaluations

Study identification (Include author, title, year of publication, journal title, pages)

Guideline topic:	Key Question No:

Checklist completed by:

Section 1: Internal validity

	In a well conducted economic study…..	**In this study this criterion is::**	
1.1	There is a defined and answerable study question	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.2	***The economic importance of the question is clear***	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	***The choice of study design is justified***	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	**All costs that are relevant from the viewpoint of the study are included and are measured and valued appropriately**	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.5	The outcome measures used to answer the study question are relevant to that purpose and are measured and valued appropriately	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.6	If discounting of future costs and outcomes is necessary, it been performed correctly	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.7	Assumptions are made explicit and a sensitivity analysis performed	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.8	The decision rule is made explicit and comparisons are made on the basis of incremental costs and outcomes	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.9	The results provide information of relevance to policy makers	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

SECTION 2: OVERALL ASSESSMENT OF THE STUDY		
2.1	Is this study an economic evaluation, or a cost analysis?	
2.2	*How well was the study done to minimise bias?* *How valid is the study?* Code +, n, or –	

Notes on the use of Methodology Checklist 6: Economic Evaluations

Section 1 identifies the study and asks a series of questions aimed at establishing the internal validity of the study under review - i.e. making sure that it has been carried out carefully, and that the results are likely to be reliable and useful. Each question covers an aspect of study design that is known to make a significant difference to the conclusions of a study.

For each question in this section you should use one of the following to indicate how well it has been addressed in the review:

- Well covered
- Adequately addressed
- Poorly addressed
- Not addressed (i.e. not mentioned, or indicates that this aspect of study design was ignored)
- Not reported (i.e. mentioned, but insufficient detail to allow assessment to be made)
- Not applicable.

1.1 There is a defined and answerable study question

As with clinical evaluations, a clearly defined question is essential to allow the user to assess how well the study has met its objectives or how relevant it is to the guideline recommendation to which the results might be applied. For an economic evaluation, the question should contain information on the alternatives under comparison, the viewpoint, and (ideally) the form of economic evaluation being used and the resulting decision rule.

1.2 The economic importance of the question is clear

Not all economic evaluations are equally relevant or important. A comparison between different drugs available to treat the same condition, for example, could influence the choice of drug and possibly the overall cost of treatment. A study of drug therapy versus psychotherapy, on the other hand, could have major implications for the range, type, and extent of resources required to deliver good quality health care in a specific area. A well conducted study will provide some information on how great an impact the results are likely to have on the overall economics of the area of health care to which it relates.

1.3 The choice of study design is justified

The design of the study can have a big impact on the results derived from it. It is therefore important that the study design is clearly identified, and its limitations made clear. Each study design has its own strengths and weaknesses and each may be appropriate under different settings.

The main types of study used for economic evaluations are:

- **Economic evaluation alongside randomised controlled trial.**

In some respects this is a good model as cost and benefit data can be collected in parallel with the clinical data, and is therefore likely to be relevant and applicable. On the other hand, a number of factors are likely to make study results unrepresentative of real practice. More resources are likely to be available in a study setting than in normal practice; patient compliance may be higher than normal; there is unlikely to be scope for economies of scale; etc. The overall result is likely to be higher costs and better outcomes in the trial than are achievable once the treatment is provided on a broader basis.

- **Before and after studies.**

A “before and after study” compares costs and outcomes before the introduction of a new therapy, and after it has been provided for some time. The major problem with this type of study design is the difficulty of attributing any changes purely to the new treatment (high risk of confounding).

- **Comparative studies.**

Two systems are compared in these studies - one with the new intervention, and one that does not have the new intervention but is similar in all other respects. This design is often used in areas where randomised trials are impractical or unethical. The main difficulty is in finding two directly comparable locations or systems and eliminating the possibility of confounding. In some studies comparisons may be made between a real location and an economic model. In all such studies use of sensitivity analysis to assess the reliability of results is essential, and such analyses are particularly important where model comparisons have been used.

- **Modelling of routine data sets.**

For major policy issues, econometric modelling based on data sets such as mortality or health service utilisation can be used to estimate the effect of changes. The general lack of suitable data sets makes this a difficult option to apply in a UK context.

- **Secondary economic evaluations.**

In these evaluations local data is applied to the results of published studies to produce economic evaluations that can be applied in the local context. The scope for applying such methods is limited by the range of published economic studies. Again, the effective use of sensitivity analysis is an essential part of a well designed study.

Whichever type of design is used, the study should make clear why it was chosen, and how any possible weaknesses were addressed.

1.4 *All costs that are relevant from the viewpoint of the study are included and are measured and valued appropriately*
This is a key aspect of study design. Any study that fails to adequately detail how cost information was obtained or estimated should not be used as evidence.

All costs relevant to the study have to be identified, measured, and valued. What constitutes "relevant costs" will depend on the viewpoint of the study. A study looking at the subject from the point of view of the health service, for example, will cover all treatment and related costs. A study taking a societal view will take into account additional costs such as lost working days.

Ideally, opportunity costs (i.e. the extent to which an opportunity to use resources for some other purpose has been given up) should be used and not purely financial costs. Costs are defined as any change (either increase or decrease) in resource use as a result of the study intervention, and measured in appropriate units.

Realistically, many studies will rely on cost data. Likely sources of such data include the financial systems of service providers, scales of charges for provision of services by the private sector, and published cost studies. All sources of cost data have weaknesses, and a well conducted study will indicate how possible uncertainties or weaknesses in the data have been addressed.

1.5 *The outcome measures used to answer the study question are relevant to that purpose and are measured and valued appropriately*

This is a key aspect of study design. Any study that fails to adequately detail how outcomes were measured and (where appropriate) valued should not be used as evidence.

All outcomes should be explicitly identified and measured, even if they are not the prime focus of the study. If, for example, a comparison of two treatment programmes showed no difference in cost effectiveness in terms of life years gained between two treatments, measurement of other factors such as long-term pain or quality of life could help choose between them.

Valuation of outcomes is only required in cost benefit analysis or other types of study where it is necessary to compare costs and outcomes in commensurate units. Even in those cases, valuation is only required where none of the options is dominant (i.e. none is clearly better and cheaper, or worse and more expensive, than the others). Methods of valuation vary considerably, and where they are used, it is essential that the valuation methods are described and associated uncertainties discussed.

1.6 *If discounting of future costs and outcomes is necessary, it been performed correctly*

In many economic studies some costs or outcomes may not arise at the time of the study, but in the future. A transplant patient, for example, may be able to resume a full life following transplant but will require lifelong drug therapy and periodic follow-up visits to hospital. These future costs and benefits must be taken into account, but should be valued at a lower level than immediate costs and benefits. This is normally done through a process of discounting at a fixed rate per annum.
Take the example of the transplant patient, and assume that following surgery he is going to be permanently reliant on drugs that currently cost £20,000 per annum. Assume also that though the actual amount paid each year remains constant, the value of this amount will decline by 6% per annum. We can now calculate how much the drug will cost in each future year, based on present day values

Year	Future value	Discount factor	Present value
0	£20,000	1	£20,000
1	£20,000	0.943	£18,860
2	£20,000	0.89	£17,800
3	£20,000	0.84	£16,800
4	£20,000	0.793	£15,860

The discount factor is calculated by working out the value of £1 less the decrease in value over the year, so in year one it is 1/1.06, in year 2 it is 0.943/1.06, and so on.

Looking at the table, it is clear that working out the cost of the drugs at a fixed rate per annum will give a very different answer to one based on the discounted rate. This is a rather simplified example, but for the purposes of study evaluation it is not necessary to evaluate such calculations in detail – just to be sure that they have been done if the interventions have long term effects, and that there is some justification for the selected discount rate.

1.7 *Assumptions are made explicit and a sensitivity analysis performed*

Economic evaluation requires assumptions to be made, but if studies are to be useful to others and comparable with other work the assumptions made must be explicit. **If a study appears to make assumptions that are not identified or explained it should not be used as evidence.**

Wherever assumptions have been made, sensitivity analyses should be carried out to see what difference variations in the assumptions would make to the final outcome. Where such analyses are not included in a study, the results should be treated with great caution.

1.8 *The decision rule is made explicit and comparisons are made on the basis of incremental costs and outcomes*

The decision rule specifies the basis on which a decision about the intervention will be made – e.g. the most cost effective option will be selected. The results of an economic evaluation are normally expressed as the additional cost per additional unit of outcome. If the results are presented in some other way, the study may not be a true economic evaluation but a form of cost study.

Note that this information provides a basis for decision making, but does not represent a decision in itself: the final decision (like the recommendations based on these studies) is likely to be influenced by other factors as well as the economic case.

1.9 *The results provide information of relevance to policy makers*
Study results should be presented clearly and concisely, in a way that makes it easy for decision makers to interpret the results correctly. Ideally, the limitations of the study should be discussed along with comments on its generalisability.
Section 2 relates to the overall assessment of the paper. It starts by asking a fundamental question about the nature of the study, and whether it is a true economic evaluation. If the paper is a cost study, it will be of little or no value as a source of evidence for grading recommendations.
The following question asks you to decide how well the study meets the quality criteria overall. This should be based on your assessment of the criteria set out in Section 1, and should use the following scale:

+	**Strong. All or most** of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.

n	If the answers to the questions pertaining to PLUS or MINUS criteria do not indicate that the report is exceptionally strong or exceptionally weak, the report should be designated with an n for neutral
-	**Weak. Few or no** criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

Considered judgement on quality of evidence

Key question:	Evidence table ref:
1. Volume of evidence *Comment here on any issues concerning the quantity of evidence available on this topic and its methodological quality.*	
2. Applicability *Comment here on the extent to which the evidence is directly applicable to chiropractors.*	
3. Generalisability *Comment here on how reasonable it is to generalise from the results of the studies used as evidence to the target population for this Best Practice document.*	
4. Consistency *Comment here on the degree of consistency demonstrated by the available of evidence. Where there are conflicting results, indicate how the group formed a judgement as to the overall direction of the evidence*	
5. Clinical impact *Comment here on the potential clinical impact that the intervention in question might have – e.g. size of patient population; magnitude of effect; relative benefit over other management options; resource implications; balance of risk and benefit.*	
6. Other factors *Indicate here any other factors that you took into account when assessing the evidence base.*	

7. Evidence statement *Please summarise the development group's synthesis of the evidence relating to this key question, taking all the above factors into account.*	
8. Recommendation *What recommendation(s) does the document development group draw from this evidence? Please indicate the grade of recommendation(s) and any dissenting opinion within the group.*	**Grade of recommendation**

Prognosis Checklist

1. Bibliographic Reference

1.0 Objectives and Hypotheses
- 1.1 Are the objectives of the study clearly stated?

2.0 Design (Methods)
- 2.1 Is the study design suitable for the objectives?
- 2.2 Who/what was studied?
- 2.3 Was a control group used if appropriate?
- 2.4 Were outcomes defined at the start of the study?
- 2.5 Was this the right sample to answer the objectives?
- 2.6 Were patients at a uniformly early stage in their disease?
- 2.7 If subgroups with important prognostic differences are already known, were there adjustments made for these prognostic factors?
- 2.8 Is the study large enough to achieve its objective? Have sample size estimates been performed?
- 2.9 Were all subjects accounted for?
- 2.10 Were all appropriate outcomes considered?
- 2.11 Has ethical approval been obtained if appropriate?

3.0 Measurement and Observation
- 3.1 Is it clear what was measured, how it was measured and what the outcomes were?
- 3.2 Was the assessment of outcomes blinded?
- 3.3 Was follow up sufficiently long and complete?
- 3.4 Are the measurements valid?
- 3.5 Are the measurements reliable?
- 3.6 Are the measurements reproducible?

4.0 Presentation of Results
- 4.1 Are the basic data adequately described?
- 4.2 Are the results presented clearly, objectively, and in sufficient detail to enable readers to make their own judgment?
- 4.3 How large are the effects within a specified time?
- 4.4 Are survival curves presented?
- 4.5 Are the results internally consistent, i.e., do the numbers add up properly?

5.0 Analysis
- 5.1 Are the data suitable for analysis?
- 5.2 How precise are the prognostic estimates?
- 5.3 Are the methods appropriate to the data?
- 5.4 Are the statistics correctly performed and interpreted?

6.0 Discussion
- 6.1 Are the results discussed in relation to existing knowledge on the subject and study objectives?
- 6.2 Is the discussion biased?

THIS CHECKLIST **USED BY NON-MUSCULOSKELETAL, WELLNESS AND SPECIAL POPULATIONS TEAM ONLY**

Methodology Checklist: Controlled clinical (non-randomized) trials, **pilot, single group and other small studies**			
Study identification (*Include author, title, year of publication, journal title, pages*)			
Guideline topic:		Key Question No:	
Checklist completed by:			
Section 1: Internal validity			
1.1	The study addresses an appropriate and clearly focused question.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.2	The assignment of subjects to treatment groups is randomized (may not be applicable for some studies, such as single group interventions)	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	Subjects and investigators are kept 'blind' about treatment allocation	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.5	The treatment and control groups are similar at the start of the trial	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.6	The only difference between groups is the treatment under investigation	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.7	All relevant outcomes are measured in a standard, valid and reliable way	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.8	What percentage of the individuals or clusters recruited into each treatment arm of the study dropped out before the study was completed?		
1.9	Where the study is carried out at more than one site, results are comparable for all sites	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
SECTION 2: OVERALL ASSESSMENT OF THE STUDY			
	How well was the study done to minimise bias? How valid is the study? Code +, n, or –		

Notes on the use of checklist for non-randomized, pilot, single group and other small studies

For each question in this section you should use one of the following to indicate how well it has been addressed in the study:

- Well covered
- Adequately addressed
- Poorly addressed
- Not addressed *(i.e. not mentioned, or indicates that this aspect was ignored)*
- Not reported *(i.e. mentioned, but insufficient detail to allow assessment to be made)*
- Not applicable.

The assignment of subjects to treatment groups was randomized

This should be noted but for these studies, absence of randomization does NOT require that the study be rejected. If the description of randomization is poor, the study should be given a lower quality rating. Non-randomized studies with more than one group should be classed as Controlled Clinical Trials.

An adequate concealment method is used

Centralized allocation, computerized allocation systems, or coded identical containers are adequate methods of concealment, and may be taken as indicators of a well conducted study.

The treatment and control groups were similar at the start of the trial

The study should report any significant differences in study groups in relation to gender mix, age, stage of disease (if appropriate), social background, ethnic origin, or comorbid conditions. These factors may be covered by inclusion and exclusion criteria, rather than being reported directly.

All relevant outcomes measured in a standard, valid and reliable way

The primary outcome measures used should be clearly stated in the study. **If the outcome measures are not stated, or the study bases its main conclusions on secondary outcomes, the study should be rejected.** Where outcome measures require any degree of subjectivity, some evidence should be provided that the measures used are reliable and have been validated prior to their use in the study.

What percentage of the individuals dropped out before the study was completed?

Conventionally, a 20% drop out rate is regarded as acceptable. Some regard should be paid to why patients dropped out, as well. It should be noted that the drop out rate may be expected to be higher in studies conducted over a long period of time. A higher drop out rate will normally lead to downgrading.

Section 2 relates to the overall assessment of the paper. It starts by rating the methodological quality of the study, based on your responses in Section 1 and using the following coding system:

+	**Strong. All or most** of the criteria have been fulfilled. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to alter.
n	If the answers to the questions pertaining to PLUS or MINUS criteria do not indicate that the report is exceptionally strong or exceptionally weak, the report should be designated with an n for neutral
-	**Weak. Few or no** criteria fulfilled. The conclusions of the study are thought likely or very likely to alter.

Critical Assessment of a Narrative Literature Review

Bibliographic Reference:
Reference Manager No.:
Authors:
Title:
Journal:

1. Introduction
   a. Is the specific purpose of the review clearly stated? Yes____ No____
   b. Were the sources and the methods for identifying the relevant sources clearly described? Yes____ No____
   c. Were the criteria for including and excluding articles clearly identified? Yes____ No____

2. Findings
   a. Was the validity of the included articles assessed objectively? Yes____ No____
   b. Were the limitations of the results of the relevant studies identified? Yes____ No____
   c. Was variation in the findings of relevant studies critically analyzed? Yes____ No____

3. Conclusions
   a. Was a clear summary of pertinent findings provided? Yes____ No____
   b. Were the authors' conclusions supported by the evidence provided? Yes____ No____
   c. Were specific directives for new research initiatives proposed? Yes____ No____

4. References
   a. Are references relevant, current and appropriate in number? Yes____ No____

Checklist for Case Series

Bibliographic Reference:
Reference Manager No.:
Authors:
Title:
Journal:

1. Is there a clear statement of the clinical importance or rationale for reporting the cases? Yes____ No____
2. Is the case series based on a representative sample selected from a relevant population? Yes____ No____
3. Are the criteria for inclusion made explicit? Yes____ No____
4. Are the cases described clearly? Yes____ No____
5. Are all pertinent patient data reported adequately? Yes____ No____
6. Were all cases included in the series at a similar point in their natural history? Yes____ No____
7. Was follow-up long enough for important events to occur? Yes____ No____
8. Were outcomes assessed using appropriate instruments which have adequate validity, reliability, & responsiveness? Yes____ No____
9. Is there a discussion of the strengths, weakness, implications and relevance of this case series to other reported cases? Yes____ No____
10. Do the authors indicate directions for future investigations or management of similar cases? Yes____ No____

EVIDENCE TABLE for Systematic reviews, RCTs, Diagnostic studies, Cohort studies, Case-Control studies, Qualitative studies, Economic studies and Case series

Notes on the use of table (most of the information can be gleaned from the publication abstract)

Study reference Primary author, ref id, year	
Study type	e.g.: RCT, Systematic review, Cohort study, Diagnostic study, Case series , etc.
Participants Incl/Excl criteria	Total number of participants and main inclusion and exclusion criteria
Interventions	Describe by group when relevant and include n
Length of follow-up from study start	In days, weeks or months (or not applicable =na depending on study design)
Results (Outcomes related to time points)	Report values for main outcome masures by group and indicate whether differences were found including statistical significance = SS, non-significant = NS. Include side effects. (Other results in a format that relates to type of study design)
Quality/Validity Score (+, n, -)	Transfer summary quality/validity score from checklist
Notes	Any major bias or problems with generalizability