Canadian Cardiovascular Society
Leadership. Knowledge. Community.
Société canadienne de cardiologie
Communauté. Connaissances. Leadership.

Framework for Application of GRADE in CCS Guideline Development

Developed June 2015 by
Sean McMurtry MD, Chair CCS Guidelines Committee
Mike McGillion RN PhD, Member CCS Guidelines Committee
Susan Oliver BSc, CAE, CCS Director Knowledge Translation
Christianna Brooks BSc, CCS KT Co-ordinator
with input from:
Nancy Santesso, RD, MLIS, PhD McMaster University
and the CCS Guidelines Committee

Date	Status	Author	Description of Change
	Created New – Version 1.0	Susan Oliver	Created new document
Nov. 2016	Edited – Version 1.1	Mairead Byrne	Updated strength of recommendation from 'conditional' to 'weak' as per Guideline Committee.
April 2020	Edited – Version 1.2	Christianna Brooks	Removed mention of position statement.

Table of Contents

Introduction 4

Framework overview 4

1. Determine scope 4

2. Develop health care questions 5

3. Conduct evidence search and screen studies 5

4. Conduct high level risk of bias assessment 5

5. Develop recommendations 6

I. Determine the strength of recommendations: 7

II. Rate the quality of evidence for recommendations 7

III. Format recommendations for CJC publication 7

IV. Document the evidence to recommendations decision: 8

V. Vote on recommendations to achieve consensus 8

Summary of framework steps and resulting documentation 9

Appendix A: Sample descriptions of evidence search methods 10

Appendix B: Risk of Bias/Quality Checklist - by factors that lower quality 11

Appendix C: Risk of Bias/Quality Assessments - by study type 14

Introduction

The CCS uses the AGREE II Instrument as an overarching framework to guide the quality and methodological transparency of our guideline development. As part of the Agree II approach, CCS adopted the GRADE Scale for rating the strength of recommendations and the quality of evidence in 2010. Since then, our writing panels have presented the recommendations in GRADE format but readers of CCS guidelines had to consider the text, recommendation grading and the references when trying to understand how the evidence related to the recommendations.

There is a growing expectation that guideline developers document their approach to systematic reviews of evidence and development of recommendations. Now it is preferred that guideline writing panels apply GRADE with more rigour and document the "evidence review to recommendation" process. CCS is trying to balance this demand for transparency with the added workload it adds for our volunteer panels. As such, CCS has developed this framework as an aid to guideline co-chairs and writing panels.

Framework overview

This document was developed to guide co-chairs and writing panels through the systematic review of evidence and the application of GRADE when developing recommendations. It is meant to be used as a framework to a more rigorous application of GRADE and is presented in five (5) high level steps with examples for documentation that can be adapted to suit the specific needs of the writing panel:

1. Determine scope
2. Develop health care questions
3. Conduct a search for evidence
4. Conduct high level risk of bias assessment
5. Develop recommendations

The goal is to have writing panels follow a systematic approach to the development of recommendations based on evidence and produce documentation that provides transparency of process to readers and stakeholders. Given the limited word count allowed in the CJC publications, the additional documentation can be provided as supplemental material either in the CJC or on the CCS website.

1. Determine scope

Although scope is somewhat determined by the CCS topic approval process, it is important for co-chairs and primary panels to agree on the scope and consider the following elements at the onset of development:

- Ensure the title reflects the scope
- Identify health intents and expected benefits or outcomes
- Identify subjects to whom the recommendations will apply
- Specify level of health care (i.e. primary, secondary, etc.) where these recommendations are supposed to be implemented
- Specify which preventive, therapeutic and diagnostic interventions will be covered and which will be not
- Specify all relevant professional groups, patients, public, etc. who are target users or beneficiaries of these guidelines and/or whose views should be sought

- Consider resources needed for the implementation of guidelines and potential barriers to implementation

2. Develop health care questions

Develop your health care questions in PICO format and decide on important outcomes. If a guideline or systematic review already exists, you can use the questions and outcomes identified and add, edit or delete questions as needed. Alternatively, you can develop new questions and select outcomes with panel members. PICO is an approach to formulating a clinical question and finding an answer in the medical literature:

Population: who are the relevant patients?

Intervention: what is the management strategy, diagnostic test or exposure that you are interested in?

Comparison: is there a control or alternative management strategy, test, or exposure?

Outcome: what are the patient-relevant consequences of the intervention?

Sample 1: Question in Pico format:

Should exercise be recommended for elderly in long term care to prevent fractures from falls?

- Population: Elderly in long term care
- Intervention: Exercise (any type) to prevent falls
- Comparison: Usual care
- Outcomes: Falls, fractures

3. Conduct evidence search and screen studies

As you undertake your evidence search and screen studies for inclusion and exclusion, it is important that co-chairs or groups leads keep a detailed record of the search strategy. If desired, CCS can obtain the assistance of a librarian to help develop the search strategy, conduct the evidence searches and perform a high level review for exclusion. A detailed description of the literature search strategy and study selection should include the following (see Appendix A for examples):

- A listing of database(s) searched with a summary of search terms used
- The specific time period covered and the date the search was done
- A summary of inclusion and exclusion criteria
- The number of studies identified and the number of studies included

4. Conduct high level risk of bias assessment

After selecting studies, to be included in support of answering the PICO questions, panel members conduct a high level review of each study to assess the risk of bias/quality and record the results in a table (see sample below).

Sample 2: Risk of Bias Assessment Table

	Study Risk of Bias Assessment See Appendix B and C for checklists							
	Study	**Design**	**Limitations**	**Inconsistencies**	**Indirectness**	**Imprecision**	**Publication**	

Relevant PICO Question	ID	Type					Bias	Overall Quality: (Very Low ⊕; Low ⊕⊕; Moderate ⊕⊕⊕; High ⊕⊕⊕⊕)
Should exercise be recommended for elderly…	Jones et al	RT	No serious limitations	No serious inconsistencies				High
	Meader et al	RT	No serious limitations	No serious inconsistencies	Serious	Serious	None	Low
	Oliver et al	RT	No serious limitations	No serious inconsistencies	Serious	Very serious	none	Very Low

There are many tools and checklists available for assessing the risk of bias/quality of studies. We have included 2 samples (Appendix B and C) that are relatively straight forward and serve as a guide to assessing the quality of a study. Appendix B is a *Checklist for Quality Assessment* (developed by Meader et al) that can be applied to any study design and addresses specific questions related to the 5 factors that lower quality. Alternatively, Appendix C includes 4 separate checklists that are appropriate for the type of study being reviewed.

As a general rule, RCTs start as high quality evidence, case control and cohort studies start as moderate quality and observational studies start as low quality. There are 5 factors that can lower quality and 3 factors that can increase quality:

5 Factors that can lower quality

1. limitations in detailed design and execution (risk of bias criteria)
2. Inconsistency (or heterogeneity)
3. Indirectness (PICO and applicability)
4. Imprecision (number of events and confidence intervals)
5. Publication bias

3 Factors that can increase quality

1. large magnitude of effect
2. all plausible residual confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed
3. dose-response gradient

5. Develop recommendations

Writing panels will develop recommendations based on review of evidence using the high level risk of bias assessment as a guide for the quality of evidence. Developing the recommendations from evidence and achieving consensus can be a lengthy and iterative process and needs to be documented for transparency. Panels should consider values and preferences and practical tips when determining the recommendation. These will bring to the surface influencing factors in the development of each recommendation.

A clear values and preferences statement identifies how those affected by the recommendation assess possible consequences and serves to clarify the priorities and value judgements underlying the decision making process. Practical tips help guide practical implementation of recommendations.

I. Determine the strength of recommendations:

When applying GRADE, the quality of evidence reflects the extent to which our confidence in an estimate of an effect is adequate to support a particular recommendation. The GRADE approach separates the quality of evidence (very low, low, moderate, or high) from the strength of recommendations (strong or weak).

CCS uses ***strong*** or ***weak*** as qualifiers for strength of recommendations. There are 4 factors to consider when determining the strength of a recommendation:

1. **Quality of evidence:** The higher the quality of evidence, the greater the probability that a strong recommendation is indicated
2. **Difference between desirable and undesirable effects:** The greater the difference between desirable and undesirable effects, the greater the probability that a strong recommendation is indicated
3. **Values and preferences:** The greater the variation or uncertainty in values and preferences, the higher the probability that a weak recommendation is indicated
4. **Cost:** The higher the cost, the lower the likelihood that a strong recommendation is indicated

II. Rate the quality of evidence for recommendations

CCS uses the words ***high***, ***moderate, low, very low*** for rating the quality of evidence of a recommendation as defined below:

High: Further research is very unlikely to change our confidence in the estimate of effect

Moderate: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate

Low: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate

Very Low: Any estimate of effect is very uncertain

III. Format recommendations for CJC publication

Recommendations need to be formatted in the CCS GRADE format and should include values and preferences and practical tips. Begin the recommendation with ***we recommend*** (where strength and quality are strong) and ***we suggest*** (where strength and quality of evidence are weak).

Sample 3: formatted recommendation with values and preferences and practical tip:

Recommendation: We recommend calculating and discussing a patient's "Cardiovascular Age" to improve the likelihood that patients will reach lipid targets and that poorly controlled hypertension will be treated. *(Strong Recommendation, High-Quality Evidence).*

Values and preferences: The primary evaluation of risk is the modified 10-year FRS. Considering the overlap in risk factors for diabetes, a simultaneous evaluation of cardiometabolic risk for diabetes might be useful to motivate lifestyle changes. It is well known that a 10-year risk does not fully account for risk in younger individuals. In these individuals, the calculation of a Cardiovascular Age has been shown to motivate subjects to

achieve risk factor targets.

Practical tip: For patients older than 75 years of age, the Framingham model is not well validated. Though clinical studies are currently under way to address this group, at this point clinical judgement is required in consultation with the patient to determine the value of pharmacotherapy. One approach is extrapolation of the modified FRS, and this approach identifies most subjects as having intermediate to high-risk based on age.

IV. Document the evidence to recommendations decision:

When developing the recommendation, it is important to document the evidence to recommendation decision process through a remarks statement or table (these do not have to be part of the published document). When mapping the evidence to recommendation decision process, consider the quality of evidence, the balance of benefits versus harms, values and preferences and the resource implications. The decision process can be documented through test explanation or a table as shown in the samples below:

Sample 4: Documenting evidence to recommendation decision using a text statement:

There is low-quality to very-low-quality evidence for the benefits and harms of cryotherapy and CKC. Although there may be fewer recurrences of CIN2+ with CKC than with cryotherapy, the harms may be greater. The resources required are also greater for CKC, including the need for operating rooms, anaesthesia, and highly trained providers or specialists. The limited data on values and preferences of women for either treatment were considered similar. This recommendation applies to women regardless of HIV status.

Sample 5: Documenting evidence to recommendation decision using a table:

Evidence-to-recommendation table

Decision domain	Judgement	Summary of reason for judgement
Quality of evidence *Is there high or moderate quality evidence?*	Yes ☐ / No ☒	There is low- to very-low-quality evidence from non-randomized studies with no independent control (leading to high risk of bias). There was also inconsistency among studies and likely selective reporting of complications.
Balance of benefits versus harms and burdens *Are you confident that the benefits outweigh the harms and burdens for the recommended strategy?*	Yes ☐ / No ☒	Residual/recurrence rates of CIN2+ are probably greater with cryotherapy resulting in higher risk of cervical cancer and related mortality compared to CKC. However, there may be fewer complications with cryotherapy. Benefits and harms may be affected by the skills of the provider. It is unclear that the benefits outweigh the harms of providing cyrotherapy over CKC when a woman is eligible for cryotherapy or CKC.
Values and preferences *Are you confident about the assumed or identified relative values and are they similar across the target population?*	Yes ☒ / No ☐	A high value was placed on the risk of complications with CKC. The panel felt that there might not be a lot of choice provided to the patient as CKC is used now only with severe cases. Moreover, professionals tend to prefer cryotherapy, which is communicated to patients. CKC is also considered major surgery compared to cryotherapy, requiring inpatient care, so it is likely patients would prefer cryotherapy.
Resource implications *Is the cost small relative to the net benefits for the recommended strategy?*	Yes ☒ / No ☐	The resources are greater for CKC than cryotherapy, and include the need for operating rooms, anaesthesia, and skilled providers.

V. Vote on recommendations to achieve consensus

Guideline recommendations are developed by consensus of the primary writing panel and it is important to document the process to show consensus was achieved but also to manage conflict of interest. CCS prefers that panels employ a survey voting process that allows all primary panel members to agree/disagree with comments/or recuse for each recommendation. For recommendations not passing with consensus on the first vote, the panel will modify the recommendations and/or addresses the comments and re-vote. Contentious recommendations may take multiple rounds of voting.

CCS strongly recommends using a survey tool as it allows all members to participate at their convenience, keeps a record of the voting results and assists with management of COI. CCS staff can assist with the setup, administration, distribution and collation of voting surveys and results.

Summary of framework steps and resulting documentation

Proper documentation provides transparency of process to readers and stakeholders and is essential to the quality of the guideline. Given the limited word count allowed in the CJC publications, the additional documentation outlined in this framework can be provided as supplemental material either in the CJC or on the CCS website. CCS staff can assist with the formatting of tables and documentation as required. The following table summarizes the steps and related documentation:

Step	Documentation options	Where published
1. Determine scope	Statement in Introduction or abstract	CJC main article
2. Develop health care questions	Pico statements or tables	Supplementary material in CJC or on CCS website
3. Conduct a search for evidence	Text or tables with search strategy details, and counts	Supplementary material in CJC or on CCS website
4. Conduct high level risk of bias assessment	Risk of bias assessment table	Supplementary material in CJC or on CCS website
5. Develop recommendations	Recommendation statements in CCS GRADE format with V&Ps and Practical Tips For each recommendation: • a text statement or table linking evidence to recommendation • a record of the voting results that achieve consensus and manage COI	CJC main article Supplementary material in CJC or on CCS website

Appendix A: Sample descriptions of evidence search methods

Example 1:

The authors searched Ovid MEDLINE, PubMed, Ovid MEDLINE in-process and other non-indexed citations, the CardioSource Clinical Trials Database (or similar database), the Cochrane Library, EMBASE, and Google scholar from 1990 through May 15, 2013. Inclusions were: (1) human subjects only; (2) articles published in English language; and (3) subjects ages 18 and over. Exclusions were: (1) no unpublished data (abstracts) unless presented at major national or international scientific meetings and cannot be older than 2 years; (2) Dissertations, books and conference proceedings are excluded.

Specific search terms used were oral contraceptives, contraceptive patch, vaginal ring, contraception, intrauterine device, menopause, postmenopausal hormone therapy, estrogen, progesterone, preeclampsia, eclampsia, gestational hypertension, …

Example 2:

The following databases were searched for prospective or retrospective cohort studies, randomized controlled trials (RCTs), and systematic reviews to answer Question 2:

- PubMed from January 1998 to December 2009
- CINAHL from January 1998 to July 2008
- EMBASE from January 1998 to July 2008
- PsycINFO from January 1998 to July 2008
- EBM (Evidence-Based Medicine) Cochrane Libraries from January 1998 to July 2008
- Biological Abstracts from January 2004 to July 2008
- Wilson Social Sciences Abstracts from January 1998 to July 2008

Appendix B: Risk of Bias/Quality Checklist - by factors that lower quality

Meader et al. Systematic Reviews 2014, 3:82 www.systematicreviewsjournal.com/content/3/1/82

Study limitations (risk of bias):

1) Was random sequence generation used (i.e. no potential for selection bias)? Yes, no, Unclear

2) Was allocation concealment used (i.e. no potential for selection bias)? Yes, no, Unclear

3) Was there blinding of participants and personnel (i.e. no potential for performance bias)? Yes, no, Unclear

4) Was there blinding of outcome assessment (i.e. no potential for detection bias)? Yes, no, Unclear

5) Was an objective outcome used? Yes, No

6) Were more than 80%[1] of participants enrolled in trials included in the analysis (i.e. no potential reporting bias)? Yes, no, Unclear

7) Were data reported consistently for the outcome of interest (i.e., no potential selective reporting)? Yes, no, Unclear

8) No other biases reported? (i.e. no potential of other bias) Yes, No

9) Did the trials end as scheduled (i.e. not stopped early)? Yes, No

Inconsistency[2]

1) Point estimates did not vary widely? Yes, No

2) To what extent did confidence intervals overlap?

- Substantial overlap
- (all confidence intervals overlap at least one of the included studies point estimate)
- Some overlap
- (confidence intervals overlap but not all overlap at least one point estimate)
- No overlap
- (At least one outlier: where the confidence interval of some of the studies do not overlap with those of most included studies)

3) Was the direction of effect consistent? Yes, No

4) What was the magnitude of statistical heterogeneity (as measured by I^2)?

- Low (e.g. I^2 <40%)

[1] 80% drop out is given as an example here a different proportion can be used depending on the context of the systematic review area

[2] Reviewers may choose to use estimates from a subgroup analysis which may explain the inconsistency but should be cautious that such an explanation of heterogeneity may be due to the play of chance

- Moderate (e.g. I^2 40-60%)
- High (e.g. I^2 >60%)

5) Was the test for heterogeneity statistically significant (p<0.1)?

- Not statistically significant
- Statistically significant

Indirectness

1) Were the populations in included studies applicable to the decision context?

- Highly applicable
- Applicable
- Poorly applicable

2) Were the interventions in the included studies applicable to the decision context?

- Highly applicable
- Applicable
- Poorly applicable

3) Was the included outcome not a surrogate outcome? Yes, No

4) Was the outcome timeframe sufficient? Sufficient, Insufficient

5) Were the conclusions based on direct comparisons? Yes, No

Imprecision

1) Was the confidence interval for the pooled estimate not consistent with benefit and harm? Yes, No

2) What is the magnitude of the median sample size?

- High (e.g. 300 participants)
- Intermediate (e.g. 100-300 participants)
- Low (e.g. <100 participants)

3) What was the magnitude of the number of included studies?

- Large (e.g. >10 studies)
- Moderate (e.g. 5-10 studies)
- Small (e.g. <5 studies)

4) Was the outcome a common event (e.g. occurs more than 1/100)?

- Yes, No, Not applicable (i.e. not a dichotomous outcome)

5) Was there no evidence of serious harm associated with treatment? Yes, No

Publication Bias (other considerations)

1) Did the authors conduct a comprehensive search? Yes, No

2) Did the authors search for grey literature? Yes, No

3) Authors did not apply restrictions to study selection on the basis of language? Yes, No

4) There was no industry influence on studies included in the review? Yes, No

5) There was no evidence of funnel plot asymmetry? Yes, No, Unclear

6) There was no discrepancy in findings between published and unpublished trials? Yes, No, Unclear

Appendix C: Risk of Bias/Quality Assessments - by study type

Reference: NIH National Heart, Lung and Blood Institute. Accessed June 17, 2015.
http://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools

Criteria for Quality Assessment of Case-Control Studies	Yes	No	Other (CD, NR, NA)*
1. Was the research question or objective in this paper clearly stated and appropriate?			
2. Was the study population clearly specified and defined?			
3. Did the authors include a sample size justification?			
4. Were controls selected or recruited from the same or similar population that gave rise to the cases (including the same timeframe)?			
5. Were the definitions, inclusion and exclusion criteria, algorithms or processes used to identify or select cases and controls valid, reliable, and implemented consistently across all study participants?			
6. Were the cases clearly defined and differentiated from controls?			
7. If less than 100 percent of eligible cases and/or controls were selected for the study, were the cases and/or controls randomly selected from those eligible?			
8. Was there use of concurrent controls?			
9. Were the investigators able to confirm that the exposure/risk occurred prior to the development of the condition or event that defined a participant as a case?			
10. Were the measures of exposure/risk clearly defined, valid, reliable, and implemented consistently (including the same time period) across all study participants?			
11. Were the assessors of exposure/risk blinded to the case or control status of participants?			
12. Were key potential confounding variables measured and adjusted statistically in the analyses? If matching was used, did the investigators account for matching during study analysis?			

Quality Rating: (Very Low ⊕; Low ⊕⊕; Moderate ⊕⊕⊕; High ⊕⊕⊕⊕)
Additional Comments/Explanation:
Rater #1:
Rater #2:

*CD = cannot determine; NA = not applicable; NR = not reported

Criteria for Quality Assessment of Controlled Intervention Studies	Yes	No	Other (CD, NR, NA)*
1. Was the study described as randomized, a randomized trial, a randomized clinical trial, or RCT?			
2. Was the method of randomization adequate (i.e., use of randomly generated assignment)?			
3. Was the treatment allocation concealed (so that assignments could not be predicted)?			
4. Were study participants and providers blinded to the treatment group assignment?			
5. Were the people assessing the outcomes blinded to the participants' group assignment?			
6. Were the groups similar at baseline on important characteristics that could affect outcomes (e.g., demographics, risk-factors, comorbid conditions)?			
7. Was the overall drop-out rate from the study at endpoint 20% or lower of the number allocated to treatment?			
8. Was the differential drop-out rate (between treatment groups) at endpoint 15 percentage points or lower?			
9. Was there high adherence to the intervention protocols for each treatment group?			
10. Were other interventions avoided or similar in the groups (e.g., similar background treatments)?			
11. Were outcomes assessed using valid and reliable measures, implemented consistently across all participants?			
12. Did the authors report that the sample size was sufficiently large to be able to detect a difference in the main outcome between groups with at least 80% power?			
13. Were outcomes reported or subgroups analyzed prespecified (i.e., identified before analyses were conducted)?			
14. Were all randomized participants analyzed in the group to which they were originally assigned (i.e., did they use an intention-to-treat analysis)?			
Quality Rating: (Very Low ⊕; Low ⊕⊕; Moderate ⊕⊕⊕; High ⊕⊕⊕⊕)			

Additional Comments/Explanation:
Rater #1:
Rater #2:

Criteria for Quality Assessment of Observational Cohort and Cross-Sectional Studies:	**Yes**	**No**	**Other (CD, NR, NA)***
1. Was the research question or objective clearly stated?			
2. Was the study population clearly specified and defined?			
3. Was the participation rate of eligible persons at least 50%?			
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?			
5. Was a sample size justification, power description, or variance and effects estimates provided?			
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8. For exposures that can vary in the amount or level, did the study examine different levels of exposure as related to the outcome (e.g., categories of exposure, or exposure measured as a continuous variable?			
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10. Was the exposure(s) assessed more than once over time?			
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
12. Were the outcome assessors blinded to the exposure status of the participants?			

13. Was loss to follow-up after baseline 20% or less?			
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcomes(s)?			
Quality Rating: (Very Low ⊕; Low ⊕⊕; Moderate ⊕⊕⊕; High ⊕⊕⊕⊕)			
Additional Comments/Explanation:			
Rater #1:			
Rater #2:			

Criteria for Quality Assessment of Systematic Reviews and Meta-Analyses	**Yes**	**No**	**Other (CD, NR, NA)***
1. Is the review based on a focused question that is adequately formulated and described?			
2. Were eligibility criteria for included and excluded studies predefined and specified?			
3. Did the literature search strategy use a comprehensive, systematic approach?			
4. Were titles, abstracts, and full-text articles dually and independently reviewed for inclusion and exclusion to minimize bias?			
5. Was the quality of each included study rated independently by two or more reviewers using a standard method to appraise its internal validity?			
6. Were the included studies listed along with important characteristics and results of each study?			
7. Was publication bias assessed?			
8. Was heterogeneity assessed? (applies only to meta-analyses)			
Quality Rating: (Very Low ⊕; Low ⊕⊕; Moderate ⊕⊕⊕; High ⊕⊕⊕⊕)			
Additional Comments/Explanation:			

Rater #1:
Rater #2:

*CD = cannot determine; NA = not applicable; NR = not reported

Reference: NIH National Heart, Lung and Blood Institute. Accessed June 17, 2015.