

**GRADE Working Group meeting
Bologna, 2-3 May, 2006.**

Present: Alessandro Liberati, Andy Oxman, David Atkins, David Tovey, Donato Papini, Francoise Cluzeau, Gunn Vist, Helena Varonen, Holger Schüneman, James Woodcock, Jan Brozek, Jane Thomas, Joanne Lord, John Williams, Katherine Jones, Mark Helfand, Merce Marzo, Nicola Magrini, Regina Kunz, Robin Harbor, Roman Jaeschke, Rossana de Palma, Signe Flottorp, Sue Hill (Wednesday), Tessa Tan Torres, Yngve Falck-Ytter, Luciana Ballini, and an Italian who spoke so fast I did not catch his name.

1. The minutes from Lyon were approved pending two corrections.

Action: Gunn

2. Judgements about trade-offs and recommendations.

After discussions in plenum, small groups and again in plenum a small group consisting of Alessandro, Andy, David, Holger, John, Mark, Roman and Sue came up with the following, which was subsequently circulated to everyone at the meeting and agreed.

Strength of recommendations

The strength of a recommendation reflects the degree of confidence that the desirable effects of adherence to a recommendation outweigh the undesirable effects. Desirable effects can include beneficial health outcomes, less burden and savings. Undesirable effects can include harms, more burden, and costs. Burdens are the demands of adhering to a recommendation that patients or caregivers (e.g. family) may dislike, such as having to undergo more frequent tests or opting for a treatment that may require a longer time for recovery.

Although the degree of confidence is a continuum, we suggest using two categories: strong and weak.

A **strong recommendation** is one for which the panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects.

A **weak recommendation** is one for which the panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but the panel is not confident about these tradeoffs. Reasons for not being confident can include:

- absence of high quality evidence
- presence of imprecise estimates of benefits or harms
- uncertainty or variation in how different individuals value the outcomes
- small benefits
- the benefits may not be worth the costs (including the costs of implementing the recommendation)

Despite the lack of a precise threshold for going from a strong to a weak recommendation, the presence of important concerns about one or more of the above factors make a weak recommendation more likely. Panels should consider all of these factors and make the reasons for their judgements explicit.

Recommendations should specify the perspective that is taken (e.g. individual patient, health care system or society) and which outcomes were considered (including which, if any, costs).

Examples of implications of a strong recommendation are:

- **For patients:** Most people in your situation would want the recommended course of action and only a small proportion would not.
- **For clinicians:** Most patients should receive the recommended course of action. Adherence to this recommendation is a reasonable measure of good quality care.
- **For policy makers:** The recommendation can be adapted as a policy in most situations. Quality initiatives could use this recommendation to measure variations in quality.

Examples of implications of a weak recommendation are:

- **For patients:** The majority of people in your situation would want the recommended course of action, but many would not.
- **For clinicians:** Be prepared to help patients to make a decision that is consistent with their own values.
- **For policy makers:** There is a need for substantial debate and involvement of stakeholders.

Judgements about the strength of recommendations.

The tables from Melbourne (explanations of the judgements) and Lyon (a framework for making the judgements) and their usefulness for making judgements about the strength of recommendations more systematic and transparent were discussed. It was agreed to keep both tables, to revise them in light of the above, and to review these at the next meeting.

Volunteers for a small group to work on this: Holger, Sue, Nicola, Andy, Gunn.

Action: Holger, Sue, Nicola, Andy, Gunn

3. Group processes as they relate to GRADE

1st step: Agree on the critical and important but not critical outcomes.

2nd step: Agree on the evidence base.

3rd step: Agree on the overall quality.

4th step: Agree on the balance of benefits, harms, burdens and cost.

It was commented that the 1st step should be agreement to use GRADE.

Roman suggested adding a final step: Agreement on recommendations.

There was agreement that this is useful, and that it is important that the processes are explicit.

4. Quality of evidence for interrupted time series analyses

Currently all observational studies start at low quality of evidence. However, interrupted time series (ITS) analyses may, at least in some circumstances, provide higher quality evidence than other observational designs (e.g. controlled before-after studies). An example of a review of the effects of reference pricing was presented where the reviewers considered this to be the case. We agreed not to change the rules at this point by having ITS analyses start at moderate quality, but to reconsider differentiating between different types of observational studies at some point in the future, if there are several examples of problems with the current approach or an empirical basis for doing this.

5. Guidance for quality criteria

Postponed. We agreed that this would be the main topic on the agenda for the meeting in Dublin.

6. Grading animal studies

We agreed to discuss this further in Dublin.

7. Consideration of costs

Joanne Lord presented some considerations why she thinks that GRADE evidence profiles may not be useful for costs and cost-effectiveness.

Up to now we have agreed to include (critical or important) disaggregated costs (resource utilisation) as outcomes and grade the quality of evidence for these similarly to other outcomes.

A small group led by Sue will further consider how best to incorporate considerations of costs and will bring a proposal with examples to the group for the next meeting in Dublin

Action: Sue, Françoise, Jo, Tessa, Roman

8. Diagnostic tests

We have discussed grading for diagnostic tests several times now and agreed how to do this, but it needs to be written up.

A small group led by Holger will write this up and present it in Dublin. We will try to involve someone from Cochrane diagnostics group.

Action: Holger, John, Jane, Roman, David A

9. Cochrane Summary of Findings (SoF) and GRADE profiler

A new version of GRADEpro is being programmed. It is more user friendly but it still contains a number of bugs. For RevMan 5.0 there will be two different software programs that will work together; i.e. GRADEpro will retrieve information from RevMan files and undertake calculations, and RevMan will store SoF tables prepared by GRADEpro as additional tables in RevMan files.

Holger demonstrated the new version of GRADEpro. This new version will be able to open previous profiles and has a screen which looks like an Evidence Profile. There will be a separate column for Sparse data rather than having this under Other considerations.

Volunteers are needed to help write/edit the help (manual) and provide examples, and to provide feedback on the installation process.

Action: All install, play and give feedback to Holger

10. Web site

Thanks to Yngve.

Translation into different languages is under way.

11. Database of evidence profiles

We have collected a number of examples of profiles from previous GRADE Working Group discussions, pilot testing of GRADE, pilot testing of the Cochrane Summary of Findings tables, and other sources. We have developed a plan for what information to include in a database using an Excel spreadsheet. However, it is time consuming to put together this database and we have decided to wait until the new version of the GRADE profiler is ready. Katharine, John, Françoise, David A, and Sue volunteered to send examples that can be added to the database.

Action: Andy, Gunn, Vigdis

12. Comparison of two systems for grading the strength of recommendation and the quality of evidence: Healthcare consumers' understanding, valuation and preferences

Holger presented the results of this study. Participants (health care consumers attending a course) understood both GRADE (using symbols) and ACCP (using numbers and letters). Symbols (GRADE arrows) were better than numbers for conveying the strength of recommendations. Letters were evaluated as being better than (the GRADE) symbols for conveying the quality of evidence, but both were well. Conclusion: Stick with current suggestion to use symbols and that those who do not want to use symbols can use numbers and letters (recognising that there may be some problems with numbers).

14. Publications and use

AHRQ EPCs are looking at different grading systems. David A reported that 13 centres will be trying out GRADE.

15. Future meetings

Spring 2007, Bilbao (Merce and Pablo)

August 2007, Toronto (before or after the GIN conference)

October 2007, Sao Paulo (before or after Cochrane Colloquium)

Spring 2008, Rome

US meeting – not decided

WHO meeting – not decided