

Minutes

GRADE working group meeting

Ottawa, October 4, 2004

Attendees: Jeff Andrews, David Atkins, Dominique Broclain, Peter Bunting, Jonathan Craig, Benjamin Djulbegovic, Martin Eccles, Yngve Falck-Ytter, Paul Glasziou, Gordon Guyatt, Sue Manley, Margaret Haugh, Mark Helfand, David Henry, Andrea Horvath, Roman Jaeschke, Katharine Jones, Jo Lau, James Mason, Ed Mills, Rafal Nizankowski, Susan Norris, Wytze Oosterhuis, Andy Oxman, Vivian Robinson, Holger Schünemann, Jane Thomas, Peter Tugwell, Katrin Uhlig, Helena Varonen, Gunn Elisabeth Vist, Joseph Watine, John Williams

1. The minutes from Helsinki, April 28 – 29, 2004 were approved.
2. Quality of Evidence for single RCTs

Katharine introduced experience using GRADE to develop WHO guidelines for malaria treatment. It was noted that systematic reviews were not available when the evidence profiles were prepared. We worked through two examples (artesunate plus sulphadoxine-pyrimethamine vs amodiaquine plus sulphadoxine-pyrimethamine and dihydroartemisinin-piperaquine vs artesunate (3 days) plus mefloquine). We discussed whether reporting bias was a concern and agreed that the quality of evidence should not be lowered because of this. It was pointed out that 14 day outcomes were not included in the evidence profiles, apparently because the guideline developers thought this was not an inappropriate outcome. It might be helpful to make that judgment explicit. It was suggested that the investigators should be contacted to obtain missing information about study limitations. We reiterated the need for information about the incidence of an outcome to make judgments about its importance. It was noted that confidence intervals should be included for the absolute effects. It was suggested that for adverse effects it might be worthwhile to obtain additional information from observational data.

We decided that for the criterion of strong association at least 2 studies and $p < 0.01$ should be fulfilled for both RCTs and observational studies.

We discussed possible indications for downgrading the quality of evidence because of publication bias, including asymmetrical funnel plots, trial registers and a possible time lag. We agreed that further guidance is needed. Paul Glasziou volunteered to draft this.

Action: Paul Glasziou

Artesunate plus sulphadoxine-pyrimethamine vs amodiaquine plus sulphadoxine-pyrimethamine.

The setting for this example is areas of low resistance in Africa (the same as the population in the RCT). A separate profile would have to be produced for high resistance areas, for which the quality of evidence would be lower because of indirectness. The quality of the study was lowered because of “randomized but no details about concealment or blinding”, the outcomes are hard outcomes based on parasites present and the limitation because of potential lack of blinding was discussed, but no change was made.

Consensus:

Clinical failure (PCR unadjusted):	Moderate quality, critical to question
Clinical failure (PCR adjusted):	Moderate quality, critical
Total failure (PCR unadjusted):	Moderate quality, critical

Total failure (PCR adjusted):	Moderate quality, critical
Serious adverse events:	Low quality, important but not critical to this question
Quality of evidence:	Moderate
Balance of benefits and harms:	Uncertain, but favor old
Balance of benefits and costs:	?
GRADE Judgement:	Don't do the new

Dihydroartemisinin-piperaquine vs artesunate (3 days) plus mefloquine

Consensus:

Clinical failure by day 56 (PCR adjusted):	Moderate quality, critical
Clinical failure by day 56 (PCR adjusted):	Moderate quality, critical
Total failure by day 56 (PCR unadjusted):	Moderate quality, critical
Total failure by day 56 (PCR adjusted):	Moderate quality, Critical
Serious adverse effects by day 56:	Moderate quality, critical

Quality of the evidence:	Moderrate
Balance of benefits and harms:	No difference
Balance of net benefits and costs:	making a change costs
GRADE judgement:	Probably don't do it

Ben suggested that the evidence profiles we have discussed in our meetings would be helpful as part of guidance. One possibility would be to include these in a library of evidence profiles in GRADEpro.

Action: Gunn Vist, Holger Schünemann, Andy Oxman

Gordon suggested adding “trials stopped early” as a criterion for lowering the quality of evidence. We agreed to discuss this further, pending further background from Gordon.

Action: Gord Guyatt

We discussed consistency of results. In addition to concerns about this criterion when there is only one trial (see below) there are issues about the heterogeneity and generalisability of results. Some of these are issues of patient population that should be specified in the setting of the question and addressed as an issue of directness. It was agreed that we need good examples to illustrate problems with the current approach before making changes.

Action: ALL

We discussed concerns that were raised at the meeting in Helsinki and previously about applying the current approach when there is only one study, and the note from David Atkins, Holger Schünemann and Andy Oxman that was circulated with the minutes. Both the second and third options in the note were considered acceptable, with a preference for the third.

The second option was to revise the definition of “inconsistent results” to include downgrading where consistency cannot be addressed due to only one study (possibly only for single centre trials).

Advantages: Minimal change to existing approach.

Disadvantages: This is probably not simple. For example, if there are two small studies with wide confidence intervals, would that be higher quality evidence than one larger study? This could result in downgrading single large trials. Exceptions would be needed for studies that were sufficiently large but this might be difficult to operationalise and might be better handled under other parameters (see below).

Holger pointed out that if we make this change this it should apply to both RCTs and observational studies.

The third option was no change the current criteria and to provide more explicit guidance for judgements about imprecision and directness in relationship to single studies.

Advantages: If it ain't broken, don't fix it. Currently, our experience is that single small RCTs get downgraded because of imprecision, limitations or directness. In this way the grade usually ends up reflecting a reasonable judgement of how certain we are, but it can be somewhat arbitrary how these criteria are used. More explicit guidance would help, including the need for caution when there is only one (or a few) small RCT(s), which can be detailed when panels make guidelines.

We decided that we need good examples. David Atkins agreed to work on this. Holger and Gunn volunteered to work with him. We discussed using a random sample of Cochrane reviews and preparing evidence profiles for the first one or two trials, which could then be considered in light of an evidence profile for the results of the meta-analysis.

Action: David Atkins, Holger Schünemann, Gunn Vist

3. GRADE Profiler (GRADEpro)

Andy presented GRADEpro, which has been developed by Holger, Jo Mosen and Andy. There was general approval for the overall design of the software.

We discussed whether to change the wording in the question from “vs” to “in preference to”. We felt that this could leave ambiguity in situations when the answer is “no”. We will reconsider the phrasing of the question during further development.

We voted whether we should have another window that requires users to complete each judgment about directness (for the population, intervention, outcome and comparison) or whether we should have a window that pops up that reminds users to consider each type of directness. The vote was equivocal.

We need to add the time frame and footnotes to where the numbers for the calculation and listing of the event rates came from. We need the additional fields for continuous outcomes (CI, type of scale, upper and lower limit and a field explaining which direction indicates better)

In addition to accuracy, patient and policy important outcomes should be included in evidence profiles that are based on test accuracy; e.g. adverse effects of the test and costs.

We reiterated the need to consider test results as being similar to surrogate outcomes; i.e. indirect evidence of patient-important outcomes. Our confidence in the quality of the evidence of test accuracy when making judgments about whether to use a test or not depends on how confident we are that there is an effective treatment available, that the test result will for some other reason have important consequences for patients (or resource use), or that the diagnostic information has inherent value to patients for some reason.

We need more diagnosis examples. Jonathan Craig agreed together with David Atkins and others to work on this.

Action: Jonathan Craig, David Atkins, Yngve Falck-Ytter Rita Horvath, Wytze Oosterhuis, John Williams

7. Cochrane summaries of findings

Paul Glasziou ran us through a questionnaire on summary of findings to be used in Cochrane reviews. We voted to include all of the information currently included in GRADE summaries of findings.

Action: Paul Glasziou, Andy Oxman

8. Challenges faced by organizations considering using GRADE

Finland: There is interest in GRADE but no changes have been made thus far. It is considered too complex, especially for beginners.

France: GRADE is complex to explain, it is difficult to choose criteria of importance and the training is a challenge. There is a need for more practical examples.

Germany: The BMJ article is being translated to German.

Poland: There is a need for evidence profiles that incorporate costs.

UK: NICE is considering change, but is concerned about the complexity.

USA: ACCP may reconsider using GRADE. AHRQ is asking some of their EPCs to comment on some of the GRADE criteria. Although the FDA will not use GRADE, it may be influenced by it.

WHO: The group preparing guidelines for malaria treatment is the first to use GRADE.

9. Publications

Gunn Vist reported that the two BioMed Central papers are nearly ready to be resubmitted.

Once we have some more examples for diagnostic tests, we should describe this in a paper..

10. Future meetings

The next meeting will be held in Rome in April or May.

Another meeting is planned for Lyon in connection to the G-I-N annual meeting in November 2005.

11. There was general agreement that this meeting was productive in spite of prior concerns about having GRADE meetings in connection with Cochrane Colloquia.