

**Minutes GRADE Working Group meeting
October 22, 2006, Dublin**

Present: Alessandro Liberati, Andy Oxman, Ben Djulbegovic, Craig Whittington, David Atkins, David Rind, David Tovey, Dominique Broclain, Eimar NicLochlainn, Gordon Guyatt, Gunn Vist, Helena Varonen, Hilda Bastian, Holger Schünemann, James Woodcock, Jan Brozek, Jeff Andrews, Jonathan Craig, Julian Higgins, Katherine Jones, Mark Helfand, Merce Marzo, Monique , Pablo Alonso Coello, Patrick Bossuyt, Paul Glasziou, Phil Alderson, Regina Kunz, Susan Norris, Sylvie Guillo, Vivian Robinson.

- 1. The minutes from Bologna** were approved.
- 2. Judgements about the strength of recommendations.**
Agreement on the table:

Factors that can weaken the strength of a recommendation	Decision	Explanation
Absence of high quality evidence	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Imprecise estimates	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Uncertainty or variation in how different individuals value the outcomes	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Small net benefits	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Uncertainty about whether the net benefits are worth the costs (including the costs of implementing the recommendation)	<input type="checkbox"/> Yes <input type="checkbox"/> No	

Frequent “yes” answers will increase the likelihood of a weak recommendation

- Strong recommendation:** the panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects.
- Weak recommendation:** the panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but is not confident.

There was some discussion regarding the 5th factor. We agreed that there is a need to work more on the wording in a small group.

Action: Gordon, David A, Holger, David R

It was also agreed to flag and come back to the potential role of the evidence for cost-effectiveness in the consideration of costs. There is still disagreement about use of cost-effectiveness models, most GRADE members prefer disaggregated costs.

3. Group processes for applying GRADE

A frequently asked question is how to apply GRADE in a panel. The modified draft (after Bologna) was circulated with the agenda.

Holger reported that group processes has been identified as possible research agenda in the WHO, although comparisons are not yet done.

It was agreed that it is suitable as background documentation and it was suggested to use examples to explain.

4. Judgements about the importance of recommendations

The strength of a recommendation is not necessarily the same as its importance. Differences in importance can be due to the magnitude of the net benefits, their impact on inequities, efficiency, total cost, severity of the problem, sustainability and the potential for improvement. This is an issue at the health systems level: the difference between strength of recommendations and importance are generally not important for patient-doctor decisions. After deciding on a strong or weak recommendation, then it is possible to give priority for other reasons to a policy. This should be done transparently.

5. Recommendations to only use an intervention in the context of research

We debated and decided at previous meetings that we do not want a middle recommendation, but in that middle area it might make sense to make a recommendation to only use an intervention in the context of research. Some organisations like NICE and others use this type recommendation. The current suggestion is to make this explicit, but then other issues need to be considered also. Those were identified in the attachment and include:

- A recommendation or suggestion to use or not to use an intervention can terminate relevant research in progress or prevent new research from beginning.
- A recommendation or suggestion to use an intervention can also increase the risk of rapid diffusion of ineffective or harmful interventions, and can result in significant irretrievable fixed costs of implementation.
- An “only in research” recommendation can provide an important boost for research that could reduce important uncertainties, and can reduce the risk of rapid diffusion of ineffective or potentially harmful interventions.

A suggestion was put forward to include a recommendation to use an intervention only in research as an option instead of a weak or strong recommendation. It was agreed that we should not do this, because such a recommendation requires additional information and considerations; i.e. weighing the advantages and disadvantages of recommending research (or participation in research) rather than the advantages and disadvantages of the intervention. Therefore, recommendations to use an intervention only in research should be considered subsequent to making a recommendation about the intervention, as a possible addendum. It is important to distinguish between recommending research and recommending using an intervention only use in research. Further research is often wanted. Panels may be uncomfortable making a recommendation because of insufficient evidence, however recommending research should not be viewed as a way of not making a recommendation about use of the intervention in clinical practice (e.g. not to use). A recommendation to use an intervention only in research implies a recommendation not to use it outside of research. This should be made explicit.

6. Judgements about the strength of evidence from animal models

Most consider evidence from animals regarding benefits as very low/ low quality, but sometimes as stronger when considering harms. Holger reported that members of the WHO avian flu guideline panel had problems considering animal research as very low quality evidence, probably because of their belief in biological models. It was suggested that in situations like this a panel could/should make recommendations for having a protocol available when the potential emergency appears (e.g. SARS, avian flu). Another example was anthrax. All the evidence is from animals (monkeys) and suggests that the vaccine is effective. There is no information from humans that the vaccine is safe or effective. Nonetheless, most people would like want the vaccine if they were exposed.

Challenges with animal studies are that they are often poorly conducted and animal models may not be a good indicators of what will happen in humans. Animal studies are often conducted in animals that are not closely related to humans. Situations when animal studies could potentially be considered as relevant evidence include when there is a lack of evidence from humans, when there are important issues of safety, and for informing judgements about resistance (where in vitro studies may also be informative).

Empirical data on how well animal studies predict results from subsequent human studies could help provide better guidance for when and how to grade animal studies (e.g. Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans. *BMJ* 2004; 328:514-7.).

7. Guidance for quality criteria (limitations, consistency, directness, sparse data, etc)

We split into small groups and reported back.

Sparse data reported by Gordon,

Judgements depend on the thresholds for patient importance and precision. Evidence needs to be precise enough to make a confident decision, which means that precision is context specific. It is challenging to define minimal important difference (MID). This is something that a guideline panel should consider, but it is outside the scope of a systematic review, because the MID depends on values. An alternative approach would be not to consider the MID in when making a judgement about precision, but to consider this when making a judgement about the balance between desirable and undesirable effects. However, it is difficult to define what “sufficiently precise” means without doing this in relationship to the MID. In this context it is a judgement about how confident are we in the estimate of effect. This is difficult to discuss or communicate to learners without considering. Guidance for judgements that are dependent on values would have advantages, but no clear proposals were put forward. The MID it is based on absolute measures, but precision is typically considered in relationship to relative effects. It was suggested that we need some examples to work through to try to move this discussion forward at the next meeting. Gordon and Pablo volunteered to work further on this.

Action: Gordon, Pablo, Holger

Limitations reported by Julian and Paul

Julian presented work related to the update of the Cochrane Handbook and how assessments of the risk of bias will be handled in Cochrane reviews. A tool (table for presenting information and judgements about the risk of bias) has been developed to improve

assessments of the risk of bias. It includes descriptions of what was done and judgements based on these, with mandatory and optional criteria:

- Sequence of allocation
- Allocation concealment
- Blinding
- Incomplete outcome data
- Selective reporting
- Other issues (early stopping)

It was suggested that guidance for making judgements about study limitations could build on this work.

Paul reported on the discussion of the small group that discussed this. Two challenges that were identified were hypercritical beginners and hypocritical experts. There was discussion about which domains or criteria deserve most weight? This should reflect empirical methodological evidence and what constitutes serious risks of bias, and could build on the work described by Julian.

Another challenge that was discussed is how to combine different criteria; e.g. do a lot of little flaws add up to fatal flaw? Do two serious flaws = a very serious flaw?

Possible solutions include: using fixed criteria (e.g. building on the work described by Julian), and training. There is a need for examples to illustrate how judgements are made and a need to be clear about the audience for guidance and examples.

Volunteers to write this up: Susan, Paul, Jonathan, Andy, Gunn

Action: Susan, Paul, Jonathan, Andy, Gunn

Directness reported by David A

Assessing directness begins with defining the question of interest using PICO. At the same time, policy makers may want an overall assessment of evidence without narrow, specified questions to begin with.

An assessment of the degree to which the available evidence differs from the questions being asked can begin by considering the following factors:

POPULATION includes target condition, patient demographics, community, geography, setting, co morbidity.

INTERVENTION includes dose, duration, timing, route of administration, complexity and reproducibility of intervention, skill and training.

COMPARATORS includes appropriateness of comparator, indirect comparison, usual versus best alternative care comparisons

OUTCOMES include intermediate or surrogate, timing, duration of follow up, reliability of reference standard, measurement, clinical importance.

It is then necessary to consider how likely it is that these factors would alter effects for specific outcomes, because of differences in baseline risk or effect modification. The overall assessment involves judgements and there is no simple rule for how to do this. Uncertainty about directness can lead to downgrading by one or two levels.

It was suggested to further develop the above list of factors and to expand this into a kind of checklist. For each factor it is important to consider both the likelihood that it is important and

the likely direction in which the estimate would change. Volunteers to write this up: David A, James W, Holger.

**Action: David A, James W,
Holger**

8. Consideration of costs

It is a challenge to find someone to work through good examples. Sue Hill and others volunteered to lead a group to work further and will have something to present at the Bilbao meeting. A proposal to AHRQ for funding for a GRADE meeting includes a document on costs as a deliverable.

9. Diagnostic tests

Holger is leading the writing up of our approach to grading evidence and recommendations for diagnostic tests. He has an EU grant for this. A very rough draft of the paper has been prepared. Janeck is now in Rome working with Holger on this. Hopefully something will be circulated by the end of the year and this can be discussed in Bilbao. HJS and ADO had arranged to meet with Patrick Bossyut to discuss collaboration between GRADE and the Cochrane diagnostic methods group.

10. Cochrane Summary of Findings and GRADE Profiler

PG; JH; HS and ADO submitted a proposal to be discussed by the Steering Group in Dublin for funding for implementing SoF tables in Cochrane reviews. It has been agreed that Cochrane reviews will include SoF tables beginning with the release of the next version of the Review Manager (RevMan) software. GRADEpro will be used to create the SoF tables and will work with RevMan. There are still some unresolved issues about the presentation of SoF tables.

Julian reported on the work that has been done using simulations to evaluate different ways of converting the results of analyses of continuous data into dichotomous presentations. None of the approaches that were evaluated were robust.

11. GRADE website

Organisations have been added.

12. Database of evidence profiles

Postponed until the new version of GRADEpro is ready. The data base is also included in a grant proposal. Ben has about 50 examples that can be included in the data base as soon as this is ready to go.

13. Use of GRADE, new developments

- Italian oncology guidelines used simplification of GRADE.
- Attempts at including GRADE in NICE.
- Am Thoracic Society has published a guide for guidelines and training is underway. There is a push for existing guidelines groups to use GRADE.
- Chronic obstructive resp disease 4th killer in the world - HJS leading an effort to merge guideline developers in respiratory disease.
- Translation of GRADE to German EBM textbook.
- Norwegian translation is published.
- Biomed public health will publish shortly about guidelines, use GRADE in guidelines.

- Different societies meeting with GRADE seminar and achieved agreement that GRADE is useful and to continue.
- Difficult to implement, no national agreement in Spain yet.
- BMJ commissioned to write about HIV in different areas and asked authors to use GRADE, only one did, looking to see how useful it was. Plans to use GRADE where think that it adds value.
- ACP endorsed GRADE.
- EB Practice centres are going to be using it, video of Gordon is being circulated.
- Antithrombotic guidelines are using it, in about 30 recommendations that will be published.
- UptoDate is concerned about the adequacy of peer review, working on interactive learning
- AHRQ is starting to use GRADE, too early for feedback.
- GRADE will be included in the Cochrane handbook for SoF tables. HJS & GHG writing a chapter.
- Some resistance in Finland, but now have one guideline with evidence summary table and are implementing the full GRADE in one.
- Kidney disease improving outcome group have endorsed GRADE, publication coming out soon, ¾ through the first guideline using GRADE.
- Series of bone guidelines, and transplant maybe possibly including GRADE in future.
- IQWiG want to use GRADE to make tables for patients. Tried on two topics which didn't work, want to try on patients.
- Developed GRADE EP locally in Florida but found difficult to disseminate.
- SBU, drug agency and health authority in Sweden have invited Andy, they want one system for Sweden.
- WHO has begun to use GRADE in some guidelines and continues to recommend it in their guidelines for guidelines.

14. Publications and applications

There is a need to update the BMJ article. Volunteer: Gordon, Jeff, Ben, Gunn, Paul, Alessandro, Regina, Holger. Holger will lead writing the paper on diagnostics. Sue will lead writing the paper on economics. Hilda wants to lead writing a paper for patients.

WHO avian influenza guideline published in Lancet Infectious Diseases and methods paper in PLoS Medicine.

Yngve has submitted a small conference proposal to AHRQ.

15. Future meetings

- *Spring 2007, Bilbo (Pablo and Merce)?*
16th GRADE meeting, workshops 17th and 18th April
Funding by regional government.
- *August 2007, Toronto (before or after the GIN Conference 23 to 25th)?*
GIN funded us in Lyon, ADO check if we get funding.
- *October 2007, Sao Paulo (before or after the Cochrane Colloquium)?*
Likely will be enough people there to make this worthwhile.
- *December 2007 or January 2008 (preferred), Washington DC*
Need to obligate the money by the end of this year.

- *Spring 2008, Rome*
Need to spend it during the spring, before May