

Draft minutes GRADE Working Group meeting 16 - 17 April 2007, Bilbao

Present: Francoise Cluzeau, Philip Alderson, John W Williams, Paul Glasziou, Yngve Falck-Ytter, Regina Kunz, David Atkins, Guyatt Gordon, Signe Flottorp, Andy Oxman, Gunn E Vist, Helena Varonen, Jan Brozek, Sue Hill, Holger Schunemann, Alessandro Liberati (Tuesday), Merce Marzo (Tuesday), Pablo Alonso, Mats Eliasson, Kristina Eklund, Susanna Axelsson, Bo Freyschuss, Mark Helfand, Hans de Beer, Béatrice Fervers, Damian Pattinson, Hilda Bastian, Brice Titio.

Regrets: Tessa Tan-Torres, Margaret Haugh, Robin Harbor, Roman Jaeschke, Nicola Magrini

- 1. The minutes from Dublin** were approved.
- 2. Discussion of plans for the BMJ series of articles & the first article: Grading recommendations, an introduction.**

First drafts of the BMJ series of articles were circulated in advance of the meeting. Few comments or suggestions had been forwarded. Those that has been forwarded had been dealt with.

Suggestions were put forward that this first article in the series is the place to highlight what is special about GRADE. We agreed to summarise this in a table.

As with any new system, there is an upfront cost when introducing change. However, Francoise and Phil (NICE) reported that this was a good investment and not more expensive than other approaches. Sue reported that for the WHO the main cost was the cost of meetings, and that using GRADE reduced costs because it is structured. However, there could be additional costs if it is necessary to commission additional systematic reviews. Gordon pointed out that guideline producers with limited resources that produce large numbers of recommendations may need to take shortcuts. He suggested that GRADE provides a conceptual framework that is very useful, even if organisations do not have the resources to prepare systematic reviews and evidence profiles for all of the recommendations that they make.

We discussed and agreed that for organisations to report that they use GRADE they should produce an evidence profile.

- 3. Second BMJ article: What is quality of evidence and why is it important to clinicians?**

A suggestion was put forward to clarify that the focus is on grading the quality of evidence for effects underlying clinical recommendations. This includes recommendations about the use of prognostic or diagnostic tests to provide patients with information, in which recommendations are based on effects of providing patients with prognostic or diagnostic information.

It was agreed to keep a brief section on eliciting peoples' values (the relative importance they attach to outcomes). There was some concern regarding the figure used to illustrate this. It will be modified. We agreed that it is best to use real examples that can be referenced as much as possible.

Considerations of indirectness are different for guidelines and systematic reviews, particularly for populations, interventions and comparators. It would be good to include examples of each type of directness (populations, interventions, comparators, outcomes and indirect comparisons).

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

After the Dublin meeting an email discussion continued regarding changing our definition of quality to take account of context in guidelines, particularly with respect to precision. The new definition was circulated and nobody had objected to it. The following definitions were agreed:

In the context of making recommendations: The quality of evidence reflects the extent to which our confidence in an estimate of the effect is adequate to support a particular recommendation.

In the context of systematic reviews: The quality of evidence reflects the extent to which we are confident that an estimate of effect is correct.

Gord had prepared a brief summary of proposed guidelines for grading down for imprecision that take account of these two different definitions. The proposed guidelines were circulated with the draft agenda. For both recommendations and systematic reviews the guidelines suggest considering both the confidence interval and the 'optimal information size' (OIS). For dichotomous events the OIS can be expressed as the number of events generated by a conventional sample size calculation specifying a particular alpha and beta error, effect size, and baseline event rate. Gordon proposed 300 events as a threshold under which evidence would be downgraded, as an alternative to calculating the OIS. Others found the brief presentation of the guidelines hard to follow and expressed concerns that a 300 event threshold would result in a dramatic increase in downgrading evidence due to imprecision. It was suggested that this threshold should not apply to rare outcomes. It was agreed to further develop the guidelines and to continue to discuss these.

Action: Gordon

5. Third BMJ paper: Judgements about the strength of recommendations

The use of terminology for 'strong' and 'weak' recommendations was discussed. Some people perceive that a 'strong' recommendation suggests that patient choice is not an issue. Our definition of a 'strong' recommendation is that "the panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects" with the implication for patients being "Most people in your situation would want the recommended course of action and only a small proportion would not". Alternative terms were considered, but we were not able to agree on terms to replace 'strong' and 'weak'. It was agreed that alternative terms should be included in the article, that guideline developers should be encouraged to use terms that are appropriate in their setting and language, and that we should continue to consider alternative terms that could replace 'strong' and 'weak'. Alternatives that we currently recommend include clinicians should/clinicians might, do it/probably do it, recommend/suggest, and using symbols. Hilda reported plans to do some testing of alternative symbols for the strength of evidence.

6. Judgements about the strength of evidence from animal models

Based on a systematic review that was circulated with the agenda (Pound 2004), we agreed that animal studies should start out as very low quality, because of indirectness. However, an example that may be an exception to this was raised: evidence of a protective effect of anthrax vaccination from animal studies in the context of little, if any prospect of obtaining evidence from studies in humans exposed to anthrax. It was agreed that this example, and others, if any can be found, warranted further consideration

Action: David

7. Good practice and “critical recommendations”

Robin suggested in an attachment to the agenda that it should be possible to categorise recommendations as ‘high clinical importance’ for situations where not following a recommendation carries a high risk of death or disability (regardless of the strength of the recommendation). We agreed we needed examples to work through to determine how and if to implement this. Françoise and Sue offered to provide examples before the next meeting.

Action: Françoise and Sue

8. Fifth BMJ paper: Incorporating considerations of resource use

The choice of comparator is important for resource use. Effect studies and resource use studies often compare a new drug or treatment with no treatment rather than with the current active drug or treatment.

Action: David offered to provide a good example of switching the comparator

The pre-eclampsia example was discussed. It was agreed it was a good illustration of the need to consider resource use relative to different populations, the same as for other outcomes.

The need to decide what perspective to take and to consider who pays complicates considerations of resource use. Mark suggested that a societal perspective should be the default rather than a health system perspective because private health plans may be insensitive to people who are sick dropping out. In settings with privately funded healthcare systems it is important to be careful regarding downstream assumptions.

Incorporating considerations of resource use will be the main focus of the Working Group meeting in Washington DC in January 2008. A deliverable for that meeting (which is funded by AHRQ) is a paper on incorporating considerations of resource use. More examples of evidence profiles that incorporate resource use are needed!

Action: All

9. Fourth BMJ paper: GRADEing the quality of evidence and strength of recommendations for diagnostic tests and strategies

It was suggested that we need to further clarify that evidence of test accuracy usually is indirect for patient important outcomes and when it is very serious, serious and not serious. It was also suggested that more consideration needs to be given to how a test will be used. A number of other issues were discussed that need clarification. It was agreed that the paper will be revised based on the discussion and sent for another round of feedback with one week to respond.

Action: Holger

We need more examples of evidence profiles for diagnostic tests.

Action: All

10. Cochrane Summary of Findings (SoF) tables

The Cochrane Collaboration has provided support for redesigning, user testing and evaluating SoF tables, which will be incorporated in Cochrane reviews beginning with the release of RevMan 5 in 2008. A draft SoF table was circulated with the agenda and discussed. Accompanying explanations that can be printed out from the table in the Cochrane Library are also being developed.

Paul suggested that we should take out confidence intervals for absolute effects. After discussing this we voted to keep these confidence intervals with only one vote remove them. We also discussed replacing mean differences with dichotomous presentations. One of Julian Higgins' students reviewed methods for doing this and the conclusion of this review was that none of the available methods was robust enough to recommend. However, review authors may still elect to do this using one of the available methods.

It was agreed that SoF tables should be limited to seven or fewer outcomes, ordered from the most important to the least important.

There was agreement that the term 'typical' in the heading for control group risk should be changed, but we did not agree on a better term.

11. GRADE profiler

Version 3.0 is currently on the GRADE website. A beta version will be ready by end of May. The RevMan 5 beta version is also expected to be ready by the end of May. The majority of those at the meeting had managed to download and install GRADEpro. All members of the Working Group were asked to help test the GRADEpro software.

Action: All

It was felt that the footnote function could be made more user-friendly.

We discussed whether it should be allowed to upgrade randomised trials; e.g. trials downgraded because of serious limitations, such as inadequate concealment of allocation, could potentially be upgraded if there was a strong association or dose response, in the same way that observational studies (which start at low) can be. It was suggested that there should be a warning in the software when this is done. The software should force /ask about the reasons for up and downgrading. Footnote use should be encouraged, but there should be the possibility of turning this forced entering off with a warning. Those profiles that do not use footnotes should include a statement something like "This evidence profile does not include footnotes with explanations of the judgements that were made."

Action: Jan and Holger

After discussion, it was agreed that it should be possible to downgrade all criteria by either one (serious) or two (very serious) levels.

Action: Jan and Holger

We discussed alternatives for copyrighting GRADEpro, which the Cochrane Collaboration has requested. Alternatives that were considered for which legal entity should hold the copyright were the Cochrane Collaboration, McMaster University, the

Norwegian Knowledge Centre for the Health Services, and individual members of the GRADE Working Group. We agreed (with one vote against) to register the copyright through McMaster, based on previous experience and convenience and to trademark the words GRADEpro, GRADEprofiler, and GRADE.

Action: Holger

12. GRADE website

It was agreed that for now it is sufficient with a single password that can be shared for the members section, but we may want to have individual passwords in the future.

It was suggested that when people download GRADEpro they should be asked to register so that they can be informed about updates and we have some data on who is using the software.

Some minor revisions were made to the listing of organisations that use or endorse the use of GRADE. No further problems with this listing were noted.

13. Database of evidence profiles

It was suggested that organisations that use GRADE could/should submit the evidence profiles they prepare to a database. However, SoF tables in Cochrane will be copyrighted, which might limit the possibility of including evidence profiles from Cochrane reviews. This will need to be explored further.

Phil raised the issue of receiving academic credit for the compilation of evidence profiles (including the possibility of registering with PubMed).

It was agreed that evidence profiles included in the database should have footnotes, although it should be possible to turn off prompt for footnotes in GRADEpro. It was suggested that we should have a searchable meta-section, structure and search terms. Helena will send the Finnish guidelines database structure for Yngve to consider.

Action: Helena

It was agreed that we should collaborate as much as possible with other organisations to develop the database. At present there are no funds available for developing the database, which limits its further development. Andy reported that SUPPORT, an EC FP6 funded project, may be able to provide evidence profiles for systematic reviews that are relevant to maternal and child health in low and middle-income countries within the next year.

14. Use of GRADE for patient information

Hilda presented some background on risk communication and understanding of patient information and work that she is planning related to the use of GRADE to produce patient information.

15. Support for WHO

Sue reported on progress at WHO and the need for support. WHO has included using GRADE in its guidelines for guidelines since 2003, but it largely has not been used in WHO guidelines until recently. In the last year three guideline groups used GRADE: avian influenza, post-partum haemorrhage, and opioid dependence. Several guidelines

currently being developed are also using GRADE. There is a need to build capacity. Plans include:

- Training at WHO headquarters in Geneva and in countries interested in adopting WHO guidelines.
- Collaboration with other groups including Nicola's and Holger's groups in Italy, Andy's group in Oslo. Sue would like to develop a list of groups internationally that could provide support to WHO guideline groups on short notice.
- Work is needed on how to develop international guidelines that can easily be adapted to specific settings and building capacity in low and middle-income settings to adapt guidelines.
- WHO guidelines often have broad target audiences including policy makers, public health professionals, clinicians and civil society. There is a need for examples of applying GRADE to the diverse range of recommendations that WHO makes, including public health and health system and policy recommendations.
- Work is underway on developing standards for different types of guidelines, including rapid responses that may need to be developed without the availability of systematic reviews.

16. Publications and applications

- Yngve reported a very successful workshop held in Freiburg.
- Hans reported that there is consideration of using GRADE in the Netherlands.
- Damian reported plans to implement GRADE in Clinical Evidence from September 2007, and their efforts to structure recommendations to fit with the current approach that is used in Clinical Evidence (using categories such as 'beneficial', but not making recommendations).
- David reported that AHRQ is informally using GRADE, but EPCs have found it difficult when a report has 80 different comparisons. They are working on ways of prioritising. They are also writing a chapter on grading evidence. The USPSTF is considering using GRADE and has planned a workshop.
- Merce reported use of GRADE for two guidelines in Spain and that physicians enjoyed this.
- Signe reported that the Norwegian Knowledge Centre for the Health Services (NKCHS) had decided to use GRADE exclusively in its HTA reports (which do not include recommendations). NKCHS supports the Norwegian Directorate for Health and Social Affairs, which develops guidelines. One guideline has been produced using GRADE so far. Several other groups in Norway are interested in using GRADE.
- Mark reported working with David. He emphasised the need for training.
- Alessandro reported that GRADE has been used in Italy for guidelines for oncology and tobacco. They have submitted a paper to an oncology journal that describes their experience using GRADE for new cancer drugs and have plans to offer more workshops. A regional process for developing sepsis guidelines is underway.
- Andy reported on the SUPPORT project that is developing summaries of systematic reviews in maternal and child health using GRADE. The same approach will be used by review centres supported by the Alliance for Health Policy and Systems Research. GRADE was also used to prepare background documents for a WHO conference on improving the use of evidence to inform health policy (IDEAhealth). NKCHS has been using GRADE to prepare press releases for new Cochrane reviews and HTA reports. A group led by Claire Glenton is working on developing plain language summaries using SoF tables.

- Pablo reported there is a national program in Spain that has postponed implementing GRADE, but are piloting its use.
- Beatrice reported that cancer guidelines in France are using GRADE for the quality of evidence, but have a different approach to recommendations. She has plans to introduce GRADE in an EC funded international oncology guidelines project.
- Katharina reported use of GRADE for the quality of evidence for two guidelines. Four organisations that have responsibility for developing guidelines are considering using GRADE.
- Bo reported that SBU has made a formal decision to use GRADE and is planning training.
- Holger reported a workshop for the Canadian Agency for Drugs and Technologies in Health (CADTH) in March. CADTH is considering using GRADE for its HTAs and is using it already for common drug reviews. In Italy he and Nicola's group have helped with evidence profiles for WHO guidelines, for instance the opioid dependence guidelines. A methods paper describing the use of GRADE for rapid guidelines for avian influenza will be published in PLoS Medicine in May and a description of the WHO Avia Flu guidelines was published in Lancet Infectious Diseases. He helped a small organization, an Evidence Based Nursing center producing guidelines for the South Tyrolian Nursing Association to use GRADE. In this field – oncology nursing care – the users found it frustrating because all that is available is very low quality evidence. The American Thoracic Society (ATS) has started to use GRADE for all its guidelines and needs training. They have created a 50% position for someone to train physicians to use GRADE for ATS guidelines. He has worked with the European Respiratory Society to accept GRADE and their executive committee will vote in June to accept GRADE for their guidelines. This is important as ATS and ERS, the two largest respiratory societies in the world, produce many guidelines together. Victor Montori has been working with Holger and Gordon on a presentation of GRADEs with Up-to-Date. A new chapter that includes the use of GRADE in SoF tables has been prepared for the Cochrane Handbook. Nancy Santesso from Ottawa is going to work in Rome to help with work on GRADEpro, SoF tables and plain language summaries. Holger and Jan are working with the Allergic Rhinitis in Asthma (ARIA) guidelines, an international WHO sponsored guideline, on implementing GRADE. This guideline will use GRADE in its next update. A methods paper will be written and published in Allergy. In addition, he negotiated with the editor of that journal a 3 part series on GRADE.
- Regina has written a chapter on GRADE for guideline developers in Germany.
- Mats reported that the social sector in Sweden will also try using GRADE for interventions.
- Phil reported that NICE has decided to use GRADE. From next year it will be used in all NICE guidelines. They are working with how to use GRADE in relationship to cost-utility models that NICE uses and for public health recommendations.
- Jan is working on the GRADEpro help file and a series of papers for an allergy journal, including one on diagnosis. In Poland there is interest in using GRADE, but guidelines are currently largely 'eminence-based'.
- Gordon renewed an appeal that everyone working with organisations that are using GRADE should ask them for permission to include them on our web pages. He reported that the ACCP antithrombotic guidelines are using GRADE and 20 recommendations have evidence profiles. The editors of Chest want to publish the evidence profiles with the guidelines. Up-to-date is moving forward with

implementing the use of GRADE. All of the experts that write for Up-to-Date will now need training. Other specialty groups are considering using GRADE.

- Brice reported that the NHA in France is planning to use GRADE.

It was suggested that we should further develop training materials and ways of responding to demands for training from the growing number of organisations that are using GRADE.

17. Future meetings

- 28 October 2007, Sao Paulo (after the Cochrane Colloquium)
- 10-11 January 2008, Washington DC (supported by AHRQ)
- 6-7 May 2008, Rome
- October 2008, Freiburg (before or after the Cochrane Colloquium)

18. Everyone expressed appreciation to Nieves, Rosa, Pablo and Merce for organising the meeting!