

## **Draft minutes GRADE Working Group meeting Edinburgh, 7 and 8 May 2009**

**Participants:** Robin Harbour, Iosief Abraha, Phil Alderson, Pablo Alonso, Jeff Andrews, Denis Belanger, Jan Brozek, Massimo Brunetti, Stephanie Chang, Françoise Cluzeau, Philipp Dahm, Hans de Beer, Yngve Falck-Ytter, Paul Glasziou, Gordon Guyatt, Mark Helfand, Brice Kitio, Jorma Komulainen, Ton Kuijpers, Regina Kunz, Gero Langer, Alessandro Liberati, Juha Liira, Nicola Magrini, Alison McLeod, Joerg Meerpohl, Alessandro Montedori, Mona Nasser, Andy Oxman, Piet Post, Silvia Pregno, Holger Schunemann, Vijay Shukla, Judith Thornton, Helena Varonen, Gunn Elisabeth Vist, Craig Whittington, John Williams

Regrets: Signe Flottorp, Geert van der Heijden

1. **The minutes from Freiburg** were approved pending additions regarding peoples work
2. Additional issues from last meeting in Freiburg
  - Diagnostic tests, this will be revisited at the next meeting. The allergy paper incorporates some of the discussed changes. User testing is planned for next 6 months and will be in collaboration with those doing user testing on Summary of Findings tables. It is expected that there will be some results in time for the Singapore meeting in October.
  - Good practice recommendations, without grading, this is now addressed in the article.
  - Grading the evidence for co-morbidities, we have not heard back from Clinical evidence about this.

### **JCE GRADE series - revisions**

There was a discussion about prognosis. In GRADE we focus on management alternative 1 versus 2 and patient important outcomes, so prognostics would fall outside. In Washington we discussed separate quality ratings for certain risk groups, and agreed that for certain populations based on extra extrapolation we could consider to create separate tables for populations with extra concern about directness. We will come back to prognostic issues at a later meeting and potentially include it in the article on Summary of Findings tables.

We agreed to consistently label the titles of this series in the negative frame since we are talking about downgrading i.e. inconsistency, indirectness,,,

The Journal of Clinical Epidemiology has agreed to publish 15+ articles in this GRADE series.

We need a draft of paper 3 (intro to rating quality of evidence) before starting to submit to the journal. Mark agreed to take this on and to have a draft ready within four months of the meeting.

The first paper in the series will list the members of GRADE working group as authors, the rest of the articles in the series will be according to contributor ship, with a low threshold of contribution for inclusion.

The stockings example is problematic because one of the authors is struck from the register due to suspicions of fraud. We agreed to find another example for the first example in the first article of the series, Paul and Regina will send the otitis example to Gordon.

**Action: Mark, Regina and**

**Paul**

We also agreed that there are important issues that need highlighting, and to use the stocking example in the article that discuss risk of bias. Discuss the apparent legitimate disagreement with footnote about the fraud and the need for complex groups

This change is also required in the Cochrane Handbook

#### **JCE article 1, Introduction – revised**

The group discussed the use of ‘good practice recommendations’ and motherhood statements, we would like this to challenge guideline producers to think about what they do and recommend. Craig W offered to send examples for use or not of GRADE, all are encouraged to send examples to Gordon of where grading is unhelpful and should not be done (unclear and not actionable recommendation), helpful and should not be graded, should be graded.

**Action: Craig, All**

The new example from Paul on otitis media will be included and methodology versus terminology will be discussed.

#### **JCE article 4, Rating the quality of evidence – risk of bias – revised**

#### **JCE article 5, Rating the quality of evidence – imprecision – revised**

We discussed a number of issues including the paper from Katharine Jones.

The group agreed to change the order both in text and Figure 3. This change will also have to be made in the Handbook and in GRADEpro.

We first look at the clinical threshold, then OIS.

Additional change: When OIS is not met then to downgrade TO when OIS is not met then consider downgrading

We will now use the example of malaria

#### **JCE article 6, Rating the quality of evidence – inconsistency – revised**

Based on discussions in Freiburg, changes were made to the article.

If there is low quality evidence from RCT, and we look for observational studies, and assuming then that the observational studies also have low quality of evidence. Currently we have used a meta-analytical approach and looked at differences in estimate.

We think this approach would not be supported by Cochrane, and suspects that some may also use this to suggest to always look for observational studies to look for consistency.

There is a continuum of quality; those who produce and compile SoF tables have to make judgements. The example on page 6 of SoF article does not follow the rules of upgrading, this example goes out.

In a situation when you have low quality RCT and low quality observational studies showing different results; there is inconsistency. Andy suggested including the example of BCG vaccine for tuberculosis

**Action: Andy**

One of the small groups discussed the importance of a priory subgroup selection versus after. A language change is suggested for Figure 11, to change from are you sure it is real or not TO presenting single estimate or separate estimates. It was also suggested to add a figure 12 presenting the pooled estimates. It was suggested to only rate down for the subgroups where there is a change in the conclusion/ recommendation. Gordon reminded us that a priory

subgroup analysis should always be explored, and to keep to a small number of planned subgroup analysis.

### **JCE article 7, Rating the quality of evidence – indirectness – revised**

A suggestion was made to organise the paper so that it introduces the three different types of indirectness before the paragraph on mechanisms.

There is doubt regarding the usefulness of surrogates, and when to downgrade one or two levels. Some uncertainty was raised about the principle in the example serological and blood pressure, are both indirect. Surrogates are almost always going to be rated down. Downgrading by -2 is perceived as harsh by many guideline panels.

Populations: We need examples that are acceptable to a broad audience – indicate degrees of indirectness, population and intervention. Downgrading by one or two levels depends on differences in populations (for example age and inclusion/exclusion criteria).

A suggestion was made to include information that animal studies are very indirect and that animal studies start at Very low quality of evidence. Applicability, including resistance trials could also be mentioned. The implication for systematic reviewers, if you have any other evidence then you will not consider animal studies. Guideline panels often do consider resistance pattern and evidence from animals for toxicity and the quality of that evidence. Different results may be due to bias as well as chance, do the different methods (with different assumptions) match. The assumptions in these studies/models may introduce bias.

### **JCE article 9, Preparing summary of findings (SoF) tables**

There is still ongoing work on how to incorporate and present prognostic uncertainty and continuous outcomes.

The group agreed to include a full GRADE evidence profile in this paper and to refer to GRADE pro. It was also suggested to include a small narrative on how to describe a SoF table and an evidence profile to a guideline panel. Suggested examples are avian influenza or pp haemorrhage.

### **How stand alone should these articles be?**

We will include a brief summary of the article in each of them and with cross referencing to the other articles.

We agreed to use ONE reference for the statistical methods, the Cochrane handbook, as far as possible.

When there is no data for an outcome: The SoF table will have an empty line with an explanation. The overall quality of evidence grading in presence of an empty line, also needs an explanation: empty, not graded. If the outcome is critical for decision making then it should be taken into account and have an impact. The absolute rate of an outcome could help inform the decision if it is critical

In SoF tables for systematic reviews, the quality column is left blank. In SoF tables for guidelines, then reconsider the criticalness of outcomes. We need examples.

**Action: All send examples to Gordon**

### **JCE article 11, Resource use – revised**

#### **3. JCE GRADE series – new**

#### **JCE article 2, Framing the question**

We discussed the crucial element of defining the population of interest. Maybe setting should be a specific issue under population, just like time is an important part of outcomes. When trying to define the scope of a review, start out broadly and plan sub group analysis. Acknowledge the variation both within systematic reviews and guidelines, there is a need to generalize before specifying the different recommendations. Paul will provide suggested wording to Gordon.

**Action: Paul**

In relation to understanding the context, how do you know it is a question that matters? Mark will provide suggested wording on this issue to Gordon.

**Action: Mark**

Article 3 should be written asap so that we can submit soon. Mark volunteered to make a draft within four months. Andy, Paul, Yngve and Jeff volunteered to make a draft of articles 12, 13, and 14. Holger volunteered to continue on article 15 on group processes. The deadline for these drafts is the next meeting in Singapore.

**Action: Mark, Andy, Paul, Yngve, Jeff & Holger**

#### **4. Alternative definitions for high, moderate and very low quality evidence**

In situations where further research is not going to happen, our definitions do not work.

An alternative was presented and discussed:

High: We are confident that the true effect lies close to that of the estimate of effect.

Moderate: The true effect is likely to be close to the estimate of effect, but there is a possibility that it is substantially different.

Low: The true effect may be substantially different from the estimate of effect.

Very low: Any estimate of effect is very uncertain.

We will continue this discussion at later dates.

#### **5. GRADE for diagnostic tests**

We had a short discussion about the importance of outcomes also for diagnostic tests. We agree that even for an accurate test, if it fails to change patient important outcomes, why would you want it implemented?

#### **6. Group processes for applying GRADE**

It was suggested that we should think of this as a manual on how to apply GRADE and not only the group processes. This would require a slight change in the wording of the title.

It was also suggested that paragraphs regarding establishing a guideline panel and prioritization of PICO does is not necessary needed in this article.

It was agreed that guideline panels could benefit from a simple and quick intro to GRADE before agreeing to use it. After agreement, it is very desirable to give training in what they need in order to understand the processes. Much of the reluctance that has been seen had been due to uncertainty about how to use it. Then they need someone competent with grade to help them. It is important to have a good quality systematic review, and then grading with GRADE is not actually so much extra work. A good systematic review should be present whatever system is used, so the additional work of obtaining a good systematic review is not actually part of GRADE. This issue of the additional/perceived additional work is a very important discussion. When there is no systematic review, it can be a big problem.

It was suggested that a section on how to choose a chair could be added. It was stressed that the chair should have understanding of GRADE and the required processes in addition to experience in leading groups.

The group should also include those potentially affected by the recommendations and technical staff (with vote or not).

The article should also mention GRADEpro with reference, discuss GRADE as symbols, conflicts of interest, presenting values and preferences, methods for developing recommendations, documentation and different models about displaying votes. It was also suggested to reemphasise the importance of making GRADE evidence profiles available and to cross reference the HARPS series.

**Action: All send feedback to Holger**

### **7. Figures for explaining and teaching GRADE**

The small group with focus on teaching discussed how there are three main audiences:

1. those new to GRADE who need convincing, with examples of using grade or not
2. those who want to understand the processes but is not going to use grade
3. those who will be using GRADE

We need a new suggestion with focus on how we are outcome centric and the importance of outcomes – for teaching example. Many guideline panels are used to focus on papers (studies rather than outcomes). We prefer a clear focus and understanding that the PICO selection is done by the panel

The picture presentation and the two-step notion using different colours were well received. We want to collect the visual material and Jeff has volunteered to compile again. Everyone in the group should send teaching material to Yngve who will post it on the web.

**Action: All & Yngve**

### **8. GRADE Profiler**

GRADEpro has up to now been available to everyone free of charge; a suggestion was made to start charging for-profit/ commercial organizations for use. It was suggested to charge about half of what RevMan charges. The group agreed that if we do, the profit would be used to further develop the profiler.

The Cochrane Collaboration has agreed to fund some of Nancy's time to develop training material, and for capacity building. There is a list of changes to the Cochrane handbook chapter to be made after this meeting.

Should there be different formats of SoF tables for different audiences? Currently we have one format and the GRADEpro team is left to consider when to consult the rest of the group. Gordon wants to be included in this group.

### **9. Database of evidence profiles**

Database: We can now tag metadata to it and searching has been improved.

### **10. GRADE website**

GRADE on google get us on first page!

The GRADE members area and GRADE community area on the web pages

### **11. Use of GRADE**

Cochrane Summary of Findings tables: Holger has spoken with Lorne B and the Cochrane Collaboration is interested in logging the profiles available on their web pages. A pilot project on providing evidence profiles to panels and an evaluation process on how useful that is, is being planned by Holger, Lorne and David Tovey.

There is a rapidly increasing demand for training, the quality of SoF tables are variable. The Cochrane editorial team is onboard and training material should be available that does not include recommendations. More trainers are needed.

SIGN: Robin reports little time and no methodological development. They are trying to create space and time and want to adopt/adapt GRADE. There was a seminar before this meeting for SIGN and other Scottish organizations. SIGN would like another GRADE workshop soon and get reassurance about the time needed to use GRADE.

USPSTF: Mark referred to and interpreted the paper p 203. A possible translation: 1. Preventive services has few randomized trials or sometimes trial are not appropriate so they use the I recommendation. 2. USPSTF identify four factors for when evidence is poor but people do it and the potential preventable burden and harms, costs and current practice are factors pertinent to decisions – when they give an I recommendation, they understand that that is what you do. 3. The authors think that GRADE does not consider these factors. They has misinterpreted GRADE to think that GRADE mean do not do when evidence is lacking,

The group discussed writing a letter to respond to the article with focus on the inaccuracies. Gordon, Holger, Alessandro, Regina, Paul, Yngve and Mark volunteered to write a response to USPSTF about this and a previous article in Annals by the same authors that also made incorrect statements/ misquoted GRADE.

**Action: Gordon, Holger, Alessandro, Regina, Paul, Yngve & Mark**

NICE clinical guidelines programme: Phil reported that they are using GRADE for the quality of evidence but that they do not use the importance column in the evidence profiles and they do not use overall grade for the quality of evidence. NICE, like many others are also struggling with presentation issues. The group suggested that it would be beneficial to all if we could work together on planning and research on presentation formats, multiple interventions and use of modelling etc. The summary of the cost effectiveness prepared by NICE was discussed in Washington and again in Rome, please share further developments.

CADTH: Denis reported that they use GRADE and that they want to make sure that they are using it properly. Evidence profiles are produced, but not published. Their researchers have had problems with GRADEpro, but will try again.

WHO: Gunn reported that WHO have through the Guidelines Review Committee included as a criteria for approval that GRADE is the method used to grade both the quality of evidence and strength of recommendation for WHO guidelines. They also recommend that each guideline panel includes someone with methodological skills. There is a list in WHO of methodologists which more GRADE working group members should sign up to.

## **12. Publications, workshops, applications**

- Regina held a workshop for the German Cochrane Centre in March. 40 people who are developing guidelines participated and there is a need for more advanced training. Systematic

review book with integration of GRADE processes. She is also considering the 7<sup>th</sup> framework EU funding call where GRADE could fit in.

- Gunn is at the WHO headquarters for six months to help with guideline development processes, methods teaching and running GRADE workshops and to guide groups that use GRADE.
- Holger is helping WHO update the avian influenza guideline that will also include recommendation about the pandemic H1N1. Critical medicine has adopted GRADE and had a workshop.
- Holger also advertised that The Am Thoracic Society has a job available for applying GRADE, it could be done by distance. McMaster is working on grade issues for clinical epidemiologist.
- John is involved with comparative effectiveness with EPCs that will soon be published in JCEpi and then the EPCs should use the methods. John is also involved in developing methods guidance on diagnostic tests.
- Jorma is grading according to GRADE and produces evidence profiles and use strength of recommendations.
- Helena is working on pilot guidelines and clinician support for implementation.
- Philipp reports that the Am Urological Society uses two fold modification of GRADE. The BMJ book will use GRADE, but it is a struggle. They have a grant to create evidence profiles for genital urology for elderly people.
- Yngve reports that he and David A are invited back to Am Gastroenterology Association to re-present GRADE. Also a liver association is interested. They all struggle with the cost issue but Yngve reminds them that it is the systematic review that is the cost driver. Yngve and Jan presented to the Swedish CDC, they were asked to focus on very low quality evidence. Yngve is also involved with the Society for Internal Medicine, Cochrane stakeholder meeting and writing a GRADE chapter for EBM Neurology.
- Vijay, next year there will be a big symposium with a special session on GRADE.
- Massimo has completed a draft on experiences of working with guideline groups, and a survey on who is using GRADE, and their experiences in addition to resources.
- Brice report that the first draft of the BMJ article is in French.
- Alessandro L is continuing to use GRADE in regional agencies in Bologna and the Cochrane Centre and for recommendation for anticancer drugs. There was an article in J Clin Oncology in 2008. Alessandro is continuing to involve more clinicians from other regions and other cancers. He held a workshop with Andy at the Cochrane Centre. A monograph is freely available on the web summarizing the GRADE methodology as from the BMJ series. A book chapter by Alessandro and Silvia in Italian (health policy) supplement about methodology. He is also involved with the National guideline program.
- Nicola reports that the national program is half way through. New cancer drugs have often just one clinical trial and GRADE is a good exercise in transparency and by end of the year 5-10 evidence summaries will be prepared. WHO is quite interested in these for the essential medicines list.
- Craig is involved with a masters program in London which will include GRADE.
- Piet reports that in the Netherlands there was recently a meeting with all guideline developing organisations and an interest in GRADE.
- Alessandro M wish to flag for next meeting: what is the relationship with clinical evidence?
- Stephanie reported that the AHRQ annual meeting had a GRADE workshop, one is using GRADE. They are having a follow up meeting to hear how other people are using GRADE. Yngve and Holger will be helping out in September. Guideline.gov has a new feature: expert commentary, introducing GRADE to policy makers with Andy, Holger and others

- Gordon is co-running a course on systematic reviews at McMaster where GRADE and evidence profiles are key to the course. They are also converting this course to a distance learning course. Gordon is also involved with Up-to-date and on ACCP guidelines with Holger.
  - Pablo is still disseminating with talks and workshops and meetings to physicians. There are already some national guidelines in print where GRADE has been used. Pablo is also involved with the Gastroenterological Association. The BMJ series is translated to Spanish for BMJ Spain.
  - Gero has started teaching GRADE in the masters program for nurses.
  - Juha reports that Occupational Health in Cochrane has an interest in guidelines.
  - Mona is evaluating how to incorporate GRADE into their evidence assessment. She is happy to share, but warns that it is all in German. She is involved with a symposium in Cochrane Bahrain and a GRADE workshop there.
  - Andy is involved in a plain language summary (PLS) trial and user testing with SoF tables to inform PLS. They compared table alone, text alone or both together. The table is amended, terms and wording consistent. Another trial is underway to test the common language and understanding. The Cochrane Centre directors meeting in Copenhagen went well.
  - Paul is updating his systematic reviews to include SoF tables. Paul is also updating the evidence based system for evidence grading that will include a link to the GRADE web pages.
  - Jan has a 3 paper series in Allergy, the first and second is in press and the third in draft form. The ARI guidelines are finished. He was also involved with a workshop for the Cochrane Collaboration in Edinburgh for editors working on their own reviews, Infection diseases in America and he has a suggestion to consider an article for wikipedia
- Holger is planning a full day workshop in Singapore

### 13. GRADE leadership

Andy will not be at the next two GRADE Working Group meetings and decided to resign as chair of the group. Holger and Gordon have agreed to co-chair the working group. We have had informal leadership; there was group consensus to continue without changes. We hope Andy will consider taking on leadership again once he returns.

### 14. Future meetings

- 15 October, Singapore
- 21 and 22 January 2010 Utrecht
- 23 October 2010 Keystone, Colorado

Alessandro and Nicola mention possibility of a meeting in Bologna spring 2011  
The Cochrane Colloquium 2011 will most likely be in Madrid, Spain.

Thank you to Robin and SIGN for hosting this meeting

Thank you to Andy for excellent leadership of the group.