Experiences and perspectives with regard to use of the GRADE approach to evaluating the certainty (quality) of evidence within the production of EUnetHTA assessments

Date: 16 October 2018
Written by: NIPHNO

Disclaimer: EUnetHTA is supported by a grant from the European Commission. The sole responsibility for the content of this document lies with the authors and neither the European Commission nor EUnetHTA are responsible for any use that may be made of the information contained therein.

Experiences and perspectives with regard to use of the GRADE approach to evaluating the quality of evidence within the production of EUnetHTA assessments

1 Introduction

One of the aims of health technology assessment is to identify, appraise and synthesize individual studies about a particular research question. GRADE (acronym for "Grading of Recommendations, Assessment, Development and Evaluations") is a method to assess the certainty (or quality) of evidence from evidence syntheses [1]. The GRADE approach has been adopted by over 100 organizations that are active in creating health technology assessments, clinical practice guidelines or systematic reviews.

Detailed guidance on GRADE is available in paper series by the British Medical Journal and by the Journal of Clinical Epidemiology[2, 3]. GRADE is also explained in the GRADE Handbook at the GRADEPro https://gdt.gradepro.org and in Cochrane Handbook [4].
EUnetHTA members have access to a video recording of a GRADE training webinar that was organised by EUnetHTA [5].

The purpose of this report is to provide an overview of the experiences and perspectives of the EUnetHTA WP4 partners regarding GRADE, and to allow an informed decision about the use of GRADE within EUnetHTA. The focus of this consultation is limited to the GRADE approach for rating the quality of evidence. The use of GRADE for grading the strength of recommendation is outside the scope of this report.
The issue was raised during the WP4 LP/CoLP face-to-face meeting in April 2018, and the decision was made to carry out this exercise.

2 Methods

To develop this report we started with an inventory of existing information resources that provide information about GRADE related experiences and perspectives within EUnetHTA. These information resources were:

- *Mapping of HTA methodologies in EU and Norway;[6]*
- *Surveys (pre/post) related to the GRADE webinar from April 2018;*
- *Survey for WP4 assessment teams for completed JA3 assessments;*
- *Results of a pilot study with the GRADEpro software which was done as a part of the OTCA07* assessment on Femtosecond laser-assisted cataract surgery (FLACS).[7]
- *The methods sections of published and planned JA 3 assessments that we screened for use of* GRADE.

In addition, we invited all the WP4 partners (n=60) to complete a survey (the EUnetHTA GRADE survey) to express their views with regard to use of the GRADE approach.

To ensure a fair presentation and discussion about GRADE, we asked the GRADE Working Group to fact-check any views (after anonymization) that were critical or that could potentially be misconceptions about GRADE. The GRADE Working Group is not a EUnetHTA working group, but an informal collaboration of people worldwide with an interest in the GRADE approach. Any feedback from the GRADE Working Group is presented as *[GRADE WG: ….].*

We summarized and presented the findings together with feedback from the GRADE Working Group at a face-to-face meeting for WP4 partners that took place in Barcelona (September 2018).

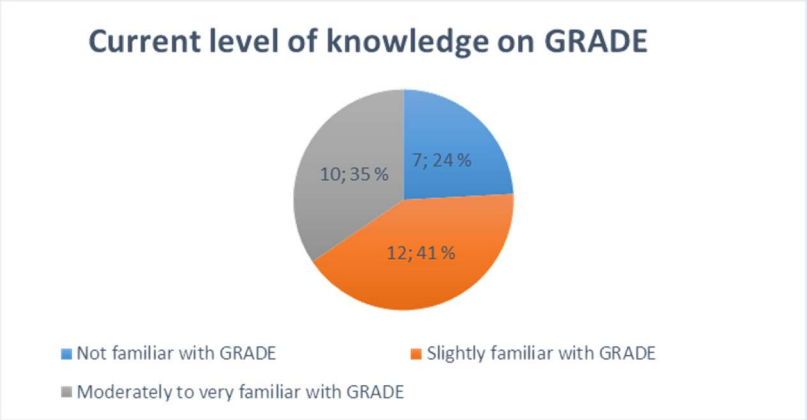
We also included a short description of GRADE related software solutions that we are aware of.

3 Results

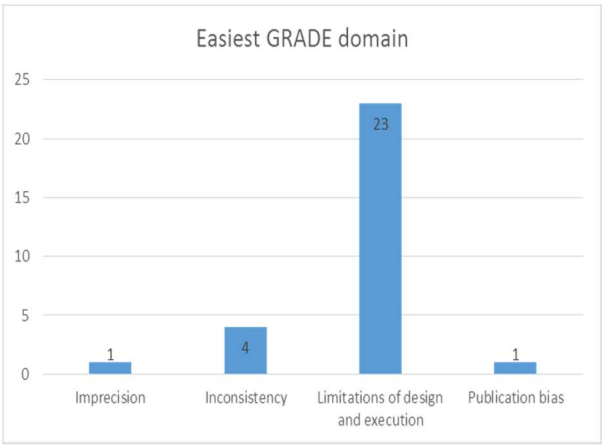
The results presented below originate from different information sources, including four surveys. The appendix provides further data about response rates to these surveys.

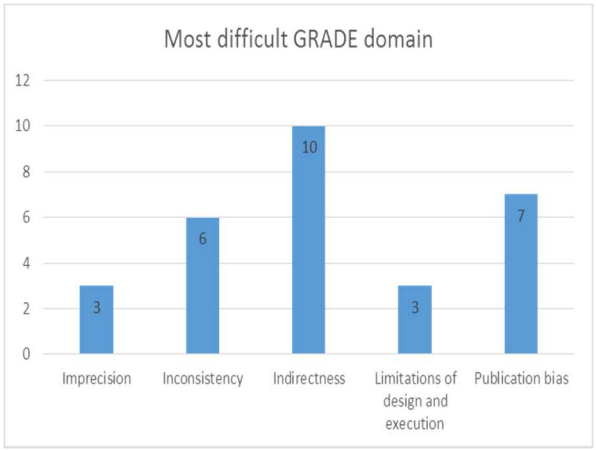
Level of GRADE related knowledge among EUnetHTA partners

Respondents (n=29) to the pre survey for the GRADE webinar indicated to have diverse levels of GRADE related knowledge. Overall, one third is not familiar with GRADE, one third is slightly familiar and one third is moderately to very familiar.



The GRADE approach contains five limitations that can lower quality: Risk of bias, Inconsistency, Indirectness, Imprecision, and Publication bias. Respondents indicated that the risk of bias domain was the easiest for them to assess, while Indirectness came out as the most difficult domain to assess.

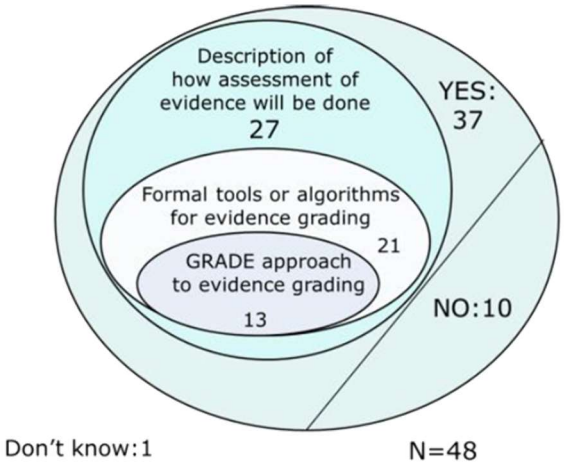




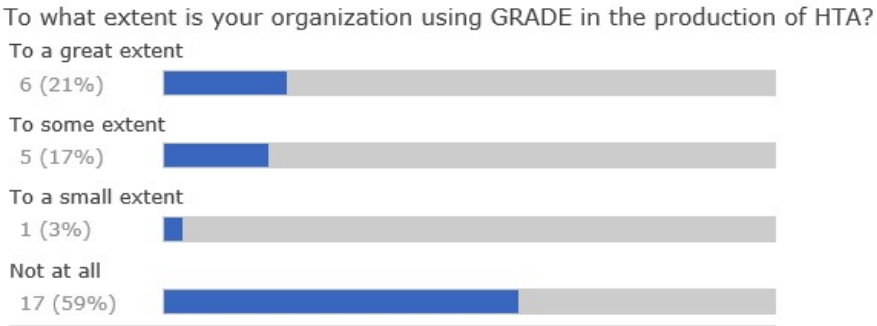
Use of GRADE within the production of HTA

The "Mapping of HTA methodologies in EU and Norway" report illustrated that GRADE has been taken up by some HTA [6]. The report states "Among the 21 institutions that include a requirement to use formal tools or algorithms for evidence grading in their pre-specified plan 13 institutions routinely use GRADE".

In total, 78 organizations received the Mapping survey and responses from 48 organizations are available for GRADE related questions.



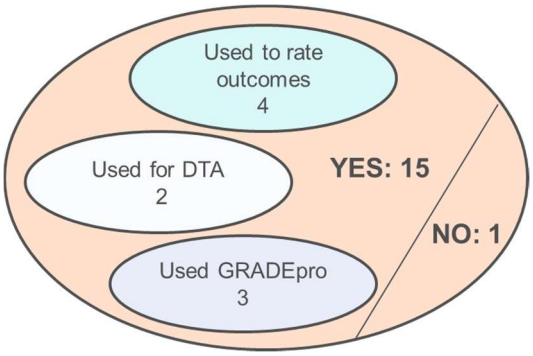
The EUnetHTA GRADE survey indicated that 12 organizations (out of 29 respondents) use GRADE to various extents.



Organizations that are only active in Pharma assessments appear to use GRADE to a lower extent than organizations that are active in Other Technologies or in both Other Technologies and Pharma assessments.

	Currently using GRADE to small/some/great extent	*Currently not using GRADE*
Other Technologies	4	5
Pharma	1	6
Both	7	6

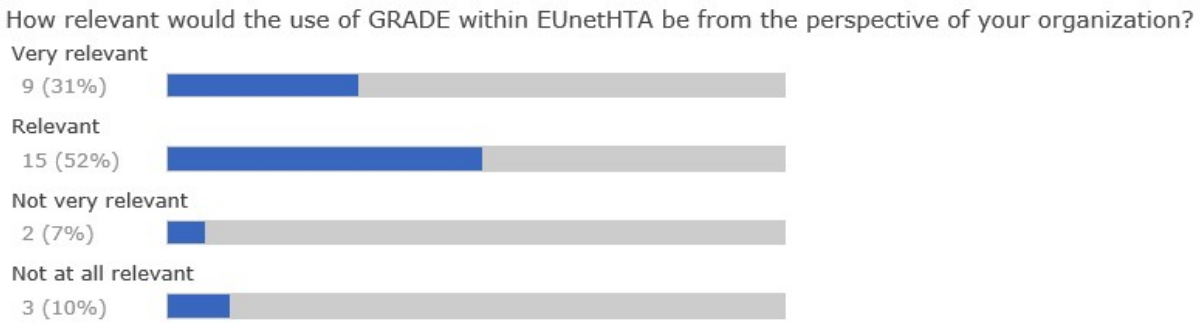
On the public EUnetHTA website, we identified 16 assessments within JA 2016-2020 for which a project plan or final report was available (as to August 2018). Screening of these reports showed that all assessments used or planned to use the GRADE approach. In one assessment, GRADE was planned in the protocol but a deviation from the protocol was necessary due to time limitations. Four assessments mentioned the use of GRADE to rate outcomes and three assessments mention the use of GRADEpro software. Two assessments used GRADE for assessments related to diagnostic test accuracy (DTA).



Relevance of using GRADE within EUnetHTA

About 80% of respondents to the EUnetHTA GRADE survey feel that using GRADE within EUnetHTA is relevant or very relevant.

Organizations that indicated that the use of GRADE within EUnetHTA is not relevant, are currently not using GRADE within their own organization.



Organizations that are active in Other technologies or in both Other Technologies and Pharma favored GRADE more, than organizations that are only active in Pharma.

	Relevant or Very relevant	*Not very relevant or Not at all relevant*
Other Technologies	9	0
Pharma	4	3
Both	11	2

Role of GRADE related software

Different GRADE related software solutions are available. This includes GRADEpro, interactive Evidence to Decision frameworks (iEtD) and MAGICapp [8-10]. These software solutions have common and unique features. Costs, potential and pros and contras for these tools have not been assessed for the purpose of this report. In addition, security issues need to be considered. For example, the use of such software within EUnetHTA might require on-site physical installation on EUnetHTA servers, which can increase the costs.

In a pilot study of GRADEpro as part of OTCA07, participants from six HTA organizations (RER, GÖG, KCE, Osteba, SESCS FUNCANIS, AQuAS) and two external experts used GRADEpro to define assessment questions, to list outcomes and to rate outcomes.

Main conclusions of the pilot study were that use of GRADEpro:

- Was a valuable and feasible option for the EUnetHTA scoping phase;
- Provided an additional means of communication and information sharing between participants in different geographical locations;
- Was a transparent process that facilitated the interpretation of evidence, making the assessment framework more robust.

Conclusions about patient or healthcare consumer involvement were that:

- No involvement of patients was attained despite several attempts;
- GRADEpro was mentioned as too technical and impersonal in relation to patient involvement. *[GRADE WG: It is not the intention that patients will do the GRADE assessment themselves.]*

Some further details are available within an abstract of this pilot study [7].

Arguments for and against the use of GRADE as perceived by EUnetHTA members

Below we present paraphrased quotes from EUnetHTA partners, which we collected through the GRADE survey and pieces of information extracted from the other available information resources. Information that is presented between square brackets summarizes the response that we received from the GRADE Working Group.

General perspectives

- "GRADE is a validated tool with explicit assessment criteria which can increase transparency of evidence assessment within EUnetHTA."
- "Arbitrary or controversial issues can be tracked and discussed within the authoring team as well as in the context of national uptake."
- "GRADE is an internationally recognized tool with increasing wide adoption."

Methodological perspectives

- "Blinding is not always possible in medical devices. Within GRADE, the authors may decide not to downgrade a study because of lacking blinding, and show this decision in a transparent way" *[GRADE WG: Blinding is irrelevant for objective outcomes (eg. mortality), but important when assessing subjective outcomes (eg. Pain). Depending on the type of outcome, authors can decide to downgrade (or not) a study because of lack of blinding (as part of the risk of bias assessment) and to report that decision transparently]*
- "GRADE could increase the risk that EUnetHTA partners become more willing to accept non-randomized evidence, even though such evidence is unable to answer effectiveness questions." *[GRADE WG: The decision to include NRS is made at the level of the HTA group. Once a group has decided to do so, GRADE appropriately evaluates the certainty of the evidence. While NRS is more prone to bias, it can help to answer effectiveness questions if no RCTs are available. NRS are essential to answer safety questions]*
- "Non statistically significant results as proof of added benefit might be accepted by EUnetHTA partners." *[GRADE WG: Reporting findings based on lack of statistically significant results is often wrongly interpreted as no effect. GRADE advocates to interpret clinical relevance (what is a clinically significant effect) together with the certainty of the evidence taking into account effect estimates and confidence intervals.]*
- "Severe publication bias might be accepted by EUnetHTA partners." *[GRADE WG: If risk of publication bias is high, one can downgrade to low, or very low if there are also concerns for other domains. This applies to all GRADE domains, including publication bias. Some GRADE users choose not to present numbers in Summary of Findings (SOF) tables when the certainty of evidence is very low, but in the end decision makers still need to make a decision.]*
- "GRADE cannot be used for all technologies (ex. public health interventions)." *[GRADE WG: GRADE can be used for all technologies also including public health interventions]*
- "GRADE does not provide explicit guidance for complex interventions." *[GRADE WG: There is some ongoing work here. GRADE CERQual provides guidance on assessment of findings from qualitative evidence syntheses]*
- "Directness (generalizability) is context dependent, and is a challenging part of GRADE if used in EUnetHTA context." *[GRADE WG: It is possible to grade differently at European level and at national level and describe it transparently through GRADE.]*
- "GRADE allows to show that there is (very) low quality evidence, which allows advocacy for better evidence."
- "The results from GRADE might be not applicable since we would normally consider lower evidence" *[GRADE WG: GRADE allows the use of all types of evidence]*

Decision maker perspectives

- "Individual agencies may choose different clinical relevance thresholds than the one selected in a EUnetHTA report and this can affect the GRADEing of evidence." *[GRADE WG: correct, clinical thresholds affect imprecision*].
- "The short and structured SOF-tables are an important tool to communicate the findings of REAs."
- "Unclear how GRADE would actually feed into decision making:"
    a. "1 trial with low certainty evidence, demonstrating significant benefit and cost-effectiveness analysis with low ICER."
    b. "How to react if the quality of evidence is low, compared to the situation when it is very low."
    c. "Usually the evidence base is very limited and GRADEing these studies does not give any useful information for decision making."

  [GRADE WG: GRADE makes it clear for decision makers when they are making decisions in uncertainty and indicates how likely it is that new studies may change the effect estimate. GRADE also formulated 5 paradigmatic situations that justify strong recommendations based on low/very low certainty of the evidence[11].]

Organisational perspectives

- "A common understanding of quality of evidence might be beneficial for a system where many countries collaborate."
- "GRADE in EUnetHTA could reduce evidence assessment work at national/local level. National uptake of EUnetHTA results might benefit."
- Training needs:
    a. «There is a lack of experience with GRADE within EUnetHTA. Learning the methodology requires time and constant updating is needed.»
    b. «There is a need for theoretical + practical sessions, targeted at beginners + advanced GRADErs. The sessions should also include examples for Medical Device assessments.»
    c. "Selection of patient relevant outcomes during PP development was difficult due to lack of consensus on what determines a patient relevant outcome"

  [GRADE WG: The GRADE WG is happy to provide further help and standardized training modules are being developed]
- «Consistent use of GRADE with good inter-rater reliability can be challenging.» *[GRADE WG: Acceptable interrater reliability has been demonstrated empirically]*
- «Experienced GRADErs are required before the start of an EUnetHTA assessment in order to obtain substantial inter-rater reliability."
- Time requirements:
    a. «Use of GRADE would increase time to undertake assessments.»
    b. "Use of GRADE is not applicable in very rapid reviews."

  [GRADE WG: Following the structured GRADE approach might also reduce time. Users need to make the trade-off between time-use and quality of decision-making]
- «SOF tables are easier to complete for authors (than initial summary table) and technology can help to create SOF tables without extra effort.»
- "Better/more stringent instructions are needed on how to present SoF tables."

Other perspectives

- "If GRADE methodology (including SOF tables and GRADEpro software) became mandatory, EUnetHTA might become dependent on GRADE."
- "GRADE would allow comparisons between EUnetHTA results and the results of other systematic reviews."
- «GRADE creates opportunities for collaboration outside HTA community and

reduced duplication of efforts.»
- «Further development of authoring technology might be needed, for example data security, usability.»

4 Discussion

The following comments were made by participants to the face-to-face meeting in Barcelona:

- LBI-HTA uses GRADE for transparency reasons and before giving recommendations.
- EUnetHTA does not want to give recommendations so the question was raised why to push GRADE. GRADE is helpful to understand differences in certainty of the evidence even if we do not proceed with giving recommendations.
- HAS does not use GRADE (often there is only 1 trial available). TLV agrees with HAS that use of GRADE in context of single arm studies is difficult. However, there is a need to describe in more detail the certainty of evidence - just the tables are not sufficient.
- ZIN points out that there might be different criteria for clinical relevance in different countries. Question was raised how to judge imprecision when you do not know the thresholds for clinical relevance.
- It was proposed that GRADE should be optional. However if it is optional, then most likely only a few would use it.
- The Senior Scientific Officer outlined that if we do not use GRADE we risk delivering a product that is not useful (i.e. that we cannot make any conclusion). We can search for a pragmatic tool that can be used instead of GRADE.
- Guidance on how to report findings benefit/harm would be helpful. It was suggested that we have dedicated people for providing support with GRADE and risk of bias at EUnetHTA. It was mentioned that further instructions are needed for summary of finding tables.

5 Conclusions

In general, findings from this report are that:

- There is a normal spread of GRADE knowledge within EUnetHTA;
- A smaller group of EUnetHTA members currently uses GRADE within their organization for the production of HTA;
- GRADE has been used in almost all JA3 assessments. This includes GRADE to assess certainty of the evidence in relation to assessments of effectiveness, safety and diagnostic test accuracy.
- Multiple software solutions are available, which have been tested within some assessments;
- A large group of EUnetHTA members find it relevant to further discuss the use of GRADE within EUnetHTA;
- Arguments pro and against have been formulated from different perspectives, which need to be further discussed;
- Further training is needed to address current knowledge gaps and misconceptions about GRADE.

The findings of this report will be presented to the WP 4 and WP6 lead partners and co- lead partners and eventually to the EUnetHTA executive board. The report will be used to support a decision of future use of GRADE within EUnetHTA assessments.

References

1. Website from the GRADE Working Group [http://www.gradeworkinggroup.org/]
2. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ: GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed)* 2008, 336(7650):924-926.
3. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A: GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of clinical epidemiology* 2011, 64(4):380-382.
4. Green S, Higgins J: Cochrane handbook for systematic reviews of interventions. In.
5. GRADE webinar [https://vimeo.com/265524107 (Password: eunethta_grade)]
6. Mapping of HTA methodologies in EU and Norway [https://ec.europa.eu/health/sites/health/files/technology_assessment/docs/2018_mapping_methodologies_en.pdf]
7. Ballini L, Bonvicini L, DeLaet C, Eisenmann A, Formoso G, Garcia-Pérez L, Giménez E, Giorgi Rossi P, Gutierrez Ibarluzea I, Ibargoyen Roteta N *et al*: Bringing GRADEpro to Health Technology Assessment: a pilot study to improve quality and transparency of Scoping and Project Plan in EUnetHTA In. Oral presentation at Cochrane Colloquium Edinburgh; 2018.
8. Website for MAGICapp [https://app.magicapp.org/]
9. Website for interactive Evidence to Decision frameworks (iEtD) [https://ietd.epistemonikos.org/#/login]
10. Website for GRADEpro [https://gradepro.org/]
11. Ignacio Neumanna, Nancy Santesso, Elie A. Akl, David M. Rind, Per Olav Vandvik, Pablo Alonso-Coello, Thomas Agoritsas, Reem A. Mustafa, Paul Elias Alexander, Holger Schünemann, Gordon H. Guyatt. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. Journal of Clinical Epidemiology 72 (2016) 45e55.

Appendix

Response rates for the different surveys that were used as information sources for this report:

- Mapping of HTA methodologies in EU and Norway: GRADE related questions answered by 48 organisations out of 78 invited EUnetHTA partners;
- GRADE webinar pre-survey: answered by 29 persons;
- GRADE webinar post-survey: answered by 17 persons;
- Survey for WP4 Assessment teams: data available from 26 survey forms;
- Survey about views on GRADE: answered by 30 organisations out of 60 invited WP4 partners.

List of organisations that responded to the survey about views on GRADE that was conducted in September 2018.

ZIN
UCSC Gemelli
LBI-HTA
NICE
SHTG
VASPVT
UTA
RER
AIFA
NIJZ
NCPE
JAZMP
FIMEA
SMC
IQWiG
SNHTA
SUKL CZ
TLV
NOMA
NSPHMPDB
DPA/MoH Malta
NIPHNO
NVD
Veneto/CRUF
SESCS/FUNCANIS
AOTMiT
AAZ
NIPN
HIQA