eunethta

EUROPEAN NETWORK FOR HEALTH TECHNOLOGY ASSESSMENT

Task Group for Common Phrases and GRADE

Partial Use of GRADE in EUnetHTA Framework

Sub-deliverable

16.11.2020

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.

Version history

Version number	Date	Finalised by	Type of document/modification	Shared with
0.1	19-10-2020	NIPHNO, ZIN	Draft subdeliverable	Task Group members
0.2	03-11-2020	NIPHNO, ZIN	Incorporated written feedback from task group members	Task Group members
Final draft	11-11-2020	NIPHNO, ZIN	Incorporated feedback e-meeting	Task Group members
Final	16-11-2020	NIPHNO, ZIN	Final version	Executive Board, Task Group

Table of contents

1 Purpose of this report

One of the objectives of the Task Group on Common Phrases and GRADE is "to recommend on the use or non-use of GRADE or other internationally adopted evidence grading system in Joint Assessments, and possibly any modifications needed". This goal extends to Collaborative Assessments as well. This report presents a proposal of standardised presentation of evidence based on partial use of GRADE and is based on the work of and discussions in the Task Group.

The proposal is put forward to the EUnetHTA Executive Board for decision making as per mandate of the Task Group. The Board will also be asked to consider whether a partner consultation is needed before the decision, or whether a partner consultation should rather be included as a recommendation in the Future Model for Collaboration White Paper. Decision by the Executive Board on use of GRADE will consequently guide the development of 'common phrases' in the Task Group (i.e. the third sub-deliverable of the Task Group).

2 Goal and use of EUnetHTA reports

The EUnetHTA Relative Effectiveness Assessments (REA's), comprising both Joint Assessments (JAs) and Collaborative Assessments (CAs), are part of a process that informs reimbursement and/or pricing decisions at a national level. At what point in this process and to what extent REA's should inform decision making is important when thinking of evidence grading systems and 'common phrases'.

A key objective for EUnetHTA REA's is that they are as informative as possible to reduce the workload of member states that implement them. However, this needs to be balanced out by the need for independent contextualization and decision making at the national level to comply with national legislation and policies.

It varies between agencies/countries, how EUnetHTA REA's are used: whether they replace national reports, or whether they are used as a basis for the creation of national reports. This influences what the different EUnetHTA partners consider to be the most appropriate product in terms of what the assessment should be comprised of.

3 Methods used by the Task Group

In 2019, WP4 Lead Partner (WP4) conducted a scoping study of existing international evidence grading systems (Appendix 1). Literature searches for reviews of grading systems, as well as input from the Task Group members, provided the evidence base for this scoping study. The focus was on mapping of strengths and weaknesses, and actual use, of the different systems. The scoping study shows that GRADE and its modified versions are more often used than other systems. Furthermore, GRADE appears to be the system that is most often evaluated favourably. Any modifications done by AHRQ ECP, NICE (in clinical guidelines only), Cochrane and others were also explored by the Task Group. However, this analysis did not consider the question if these grading systems are fit for purpose in the EUnetHTA decision making context (i.e. avoiding interference with national decision making in a multinational environment).

Based on the results of the GRADE Survey conducted by WP4 in 2018 (Appendix 2) and the scoping study above, the Task Group members discussed and explored in depth issues related to GRADE, e.g. pros and cons, feasibility, possible misconceptions and any possible requirements for modifications. A GRADE face-to-face meeting was organized in Diemen at March 2, 2020, where it was discussed if and how GRADE could be applied precisely respecting the needs of the different EUnetHTA partners. An online follow-up workshop on GRADE, focusing on a concreate example of grading, was organized on September 11, 2020.

A summary of the requirements for partial use are presented in this framework paper.

4 Core principles and prerequisites to consider for partial use of GRADE in EUnetHTA context

Based on the discussions during the workshops, it became clear that a set of conditions need to be set out for GRADE to be a beneficial and feasible approach to assess the certainty of the evidence in REA's. A number of decisions made when using GRADE were considered part of the appraisal process rather than part of the assessment by various partners. In addition, using GRADE includes methodological guidelines that might not be fully aligned with EUnetHTA guidelines.

Therefore, the Task Group concluded that GRADE needs to be applied context-independently, without *overall* conclusions on quality or certainty of evidence, and by that leave the flexibility and adaptability to be modified locally to reflect the national contexts. By refraining from overall conclusions we mean that we will not make judgements per outcome (taking into account all GRADE domains), and neither across outcomes. Secondly, while the domains of GRADE are proposed to be used for a complete, transparent and systematic assessment, EUnetHTA methodology determines the actual application.

Core principles and prerequisites discussed are listed below.

4.1 Core principles

1. REA's are part of an iterative process. In that, it is not the starting product nor the end product. The ultimate goal is to inform decision-making, but more intermediate steps/products/agencies may exist at the national level for this purpose.
2. In informing decision-making, REAs should do so to the greatest extent possible without prejudicing it. It must not interfere with national decision making processes.
3. Quality of evidence and the uncertainty thereof are crucial elements informing decision-making. However, their interpretation is partially context-dependent.
4. Assessing certainty of evidence in a REA needs to be done in a context-independent manner, allowing later (national) steps in the iterative process to translate those judgements in a context-dependent manner. Factual context-dependent information that facilitates national adaptation should be provided.
5. Assessing evidence requires that judgements are made and reported explicitly. While judgements reflecting the certainty of evidence are evidently valuable for implementation purposes, both for agencies that use GRADE as well as agencies that do not use GRADE, it is critical that judgements can be differently assessed and weighed nationally depending on national criteria. It also implies that some judgements could be irrelevant locally and therefore can be ignored in later iterative processes. An example can be an irrelevant PICO or elements of the PICO.
6. The partial use of GRADE should follow methodological guidelines from EUnetHTA. These include the domains used by GRADE, but also, for example, inclusion and exclusion criteria of studies, risk of bias assessment, indirect comparisons, outcome selection and reporting (including surrogacy and safety outcomes), handling of subgroups, presentation of results and the outcomes of the PICO subgroup. Methodological guidelines should be updated to include clear guidance on such topics.
7. Fundamental properties of GRADE are presentation of results on the outcome level and transparency on judgements.

4.2 Prerequisites

In order to have the flexibility nationally to use the REA's in a way that fits national decision-making, judgements in a EUnetHTA REA's should be made from a context-independent perspective. To facilitate national decision-making, factual context-dependent information should also be reported.

1. A disclaimer describing the scope of a EUnetHTA REA with regard to the iterative process should be created and included in the REA's. Please see chapter 5.3. for the suggestion of the Task Group on a such disclaimer.
2. No overall judgement on the certainty of evidence shall be given, By refraining from overall conclusions we mean that we will not make judgements per outcome (taking into account all GRADE domains), and neither across outcomes.
3. There should be no balancing of favourable and unfavourable effects.
4. No ranking of outcomes in terms of importance or otherwise shall be given.
5. Surrogate endpoints should be presented as measured, no assignment of surrogate effects to clinical endpoints should be performed.
6. GRADE is designed to be transparent. Therefore, all judgements within the different GRADE domains (both positive and negative) should be made transparent.
7. The Task Group focused on GRADE as an assessment tool with incorporation of existing EUnetHTA methodology. The Task Group observed that EUnetHTA methodological guidelines are currently lacking certain specific and important topics. These include: which study designs would be the most suitable to be included in REA's, how to deal with outcomes that were planned in a study but were not included in the PICO, how to deal with statistical robustness including presentation of p-values. Addressing those topics was considered out of the scope of this task group. Further development of methodological guidelines is therefore needed.

5 Proposal on how to partially use GRADE in EUnetHTA context

5.1 Basic principles on how to present the certainty of the evidence

The Task Group proposes the following basic principles:

1. We will evaluate and present aspects related to the (un-)certainty of the evidence by partial use of GRADE domains and the GRADE evidence table (Table 1).
2. We will include all relevant information in using a clear presentation in the table, not in footnotes
3. We will not downgrade or upgrade the evidence (i.e. per GRADE domain), or provide any overall judgement of the certainty of the evidence per outcome taking into account all GRADE domains (i.e. high, medium, low, very low).
4. More guidance on how to calculate OIS and where to find MIDs, is needed. Also on how this can be collected and presented in a useful way to have more explicit information/clarification on OIS and MIDs included in EUnetHTA reports. Standard formulations for the latter could be drafted.

Table 1. Example evidence table structure

Outcome	Design	Factors that may affect certainty of evidence					Number of patients		Effect estimate
		Risk of bias	Indirectness	Incon-sistency	Imprecision	Other	Intervention A	Intervention B	
Outcome 1									
Outcome 2									
Outcome *n*									

5.2 Suggested presentation of the evidence table and the different GRADE domains

- **Outcomes**

Here we enter the name of the Outcome, e.g. "Mortality" or "SAE". Outcomes are presented and assessed according to how they were measured in the studies even if they are surrogate outcomes, e.g. PFS. In case of post hoc analyses, this information can be added in this field as well.

- **Design**

Here we present the type of the study design, e.g. "randomized trial" or "observational study" (or more specifically: e.g. interrupted time series).

- **Risk of bias**

Here we enter the overall conclusion of the Risk of bias (RoB) table (last column in EUnetHTA RoB table), and any explanations for medium or high risk of bias. RoB should be assessed using the available EUnetHTA guidelines and templates.

NB: GRADE has set out guidelines on the assessment of the certainty of evidence of observational trials. Whether or not observational trials should or should not be included for assessment, and which tool should then be used to assess risk of bias in observational studies, is outside the scope of this Task Group.

- **Indirectness**

Here we assess how close the studied population, intervention, control and outcome come to the desired PICO(s). Indirectness may occur because of misfit of evidence with the population of interest, surrogate endpoints indirect treatment comparisons, or other forms of indirectness, such as dosage misalignments on the intervention level. Indirectness due to surrogacy will be 'flagged' in the assessment.

Indirectness in GRADE deals with comparing the characteristics of the included studies with the PICO(s), which for PT is prepared based on the PICO survey. This is not to be exchanged with the judgement of the directness of PICO(s), which is related to the national context and is done at the national level.

The Task Group proposes the following:

1. We use the Indirectness Tool in an adjusted format (Table 2).
2. We provide separate presentations of the identified studies / study characteristics for the different PICOs per outcome.
3. We provide a factual statement on whether one has identified any deviations - and if yes, which deviations. We do not provide more extended judgements: "yes", "probably yes", "probably no", "no". See Table 2 for an example.
4. We provide a narrative summary of the deviations in the evidence table (Table 3).
5. We will not discuss the relevance of the provided deviations in the EUnetHTA report. This is done at the national level.

Table 2. Adjusted Indirectness Tool with an example from PTJA06 (polatuzumab)

Outcome: Mortality

Domain (original question asked	Description	Assessment PICO2
Population:	Only age <65 y are included while PICO2 states all ages should be included	Indirectness issues are flagged
Intervention: Polatuzumab+BR	Dosage and use according to I in PICO2	
Comparator: BR	Dosage and use according to C in PICO2	
Direct comparison	I (Polatuzumab+BR) of PICO2 was directly compared to C (BR)	
Outcome:Mortality	Only all-cause mortality was studied not disease-related mortality as was stated in PICO2	

Table 3: Completed indirectness domain in the evidence table with an example from PTJA06 (polatuzumab) for PICO2

Outcome	Design	Factors that may affect certainty of evidence					Number of patients		Effect estimate
		Risk of bias	Indirectness	Incon-sistency	Imprecision	Other	Polatuzumab + BR	BR	
Outcome 1			- Only patients aged <65 were included, while PICO2 states all ages should be included. - The study compares with BR only, while the PICO lists (X) as relevant comparators. - DFS was included as an outcome in addition to OS, but was not studied in the trial.						

- **Inconsistency**

Here we assess the (unexplained) variation in treatment effects between studies. Statistical variation measures, like I^2 come with every meta-analytic tool, such as RevMan and could be supportive. Inconsistency in the form of heterogeneity between studies is well documented in EUnetHTA[1] and those guidelines can be followed when addressing this domain. In case of only 1 trial for a certain outcome, EUnetHTA does not assume inconsistency . See Table 4 for two examples.

Table 4. Completed inconsistency domain in the evidence table with an example from PTJA06 (polatuzumab) and from OTCA07 (FLACS)

Outcome	Design	Factors that may affect certainty of evidence					Number of patients		Effect estimate
		Risk of bias	Indirectness	Inconsistency	Imprecision	Other	Polatuzumab + BR	BR	
Outcome 1				1 study					
Outcome 2				Results of the four trials are inconsistent (results from one of four trials favouring FLACS, while results from other three studies showing no difference between study arms, with I^2=87%)					

[1] https://eunethta.eu/wp-content/uploads/2018/03/Direct_comparators_comparisons.pdf; see e.g. 2.6.1 and recommendation 4 (p6)

- **Imprecision**

Background: assessing imprecision is to a very large extent a contextual matter and may, among other elements, involve:

- statistical considerations, i.e. whether or not the effect is statistically significant given a certain α-level
- minimal (clinically) important differences (M(C)IDs), that reflect the minimal change on a certain outcome that is meaningful for the patient. For some outcomes one (or more) M(C)ID-values are published by renowned institutes (e.g. ESMO-MCBS[2] in oncology for outcomes related solid tumours), or may be published on (inter)national platforms (e.g. ICHOM).
- calculations of optimal information size (OIS).

GRADE recommends to use these three elements to further assess imprecision. It can be considered in EUnetHTA to use this as well in cases where, despite a statistical significant result, uncertainty occurs caused for example by small sample sizes and low event rates. If the OIS is used, the calculation/estimation of the required samples size and the provenance of reported M(C)ID values should be transparently reported with the judgement to allow national decision making bodies to decide if they want to follow this approach.

In the imprecision domain, considerations for problems arising from multiple testing (type 1 error) can be addressed in relation to the certainty of the evidence, especially if not appropriately adjusted.

Given the above, the Task Group proposes the following:

1. To present the following information:
    - effect estimate with confidence interval and p-value (in the same font size)
    - optionally present optimal information size (OIS) and minimal important differences (M(C)IDs) (multiple if available) from the literature etc. as long as there is no judgement following (if these approaches are used, detailed information on the provenance need to be included in the assessment report). No judgements or further interpretation of a given M(C)ID is necessary, since this is considered a local contextualization step.
2. To gather knowledge of relevant M(C)IDs from partners during the scoping phase. Additional time burden for the assessment team should be avoided.
3. To present effect estimates, but not state whether an intervention is effective or not (i.e. no conclusions).
4. To add information about pre-defined/post-hoc analyses, primary/secondary or exploratory outcomes etc. where considered relevant.

[2] https://www.esmo.org/guidelines/esmo-mcbs

Table 5. Completed imprecision domain in the evidence table with an example from PTJA06 (polatuzumab; PICO 1a)

Outcome	Design	Factors that may affect certainty of evidence					Number of patients		Effect estimate
		Risk of bias	Indirectness	Incon-sistency	Imprecision	Other	Polatuzumab + BR	BR	
Overall survival					- The effect estimate is statistically non-significant. - The number of patients is very small causing the Optimal Information Size not being met (see calculation in footnote). - Hypothetical European Clinical Organization (HECO) has published an M(C)ID: HR<0.70 for this indication.		11	12	HR 0.29 (95% CI: 0.05 to 1.64) p=…

- **Other**

Here we report other factors that might affect the confidence. For RCTs, it's about publication bias if it can be assessed through a funnel plot, where the size of the effect is plotted against the included study sizes. However, these plots are only assessable with about 4-5 studies or more[3]. If applicable, funnel plots should be published in EUnetHTA reports.

Selective reporting of outcomes is assessed as part of RoB. To avoid double count, it should not be included under the "Other" domain.

Respecting paragraph 5.1, point 4; in case observational trials are included, relevant factors could additionally include 1. Whether a large magnitude of effect exists, 2. when there is a dose-response gradient, and 3. when all plausible confounders or other biases increase confidence in the estimated effect.[4]

The Task Group proposes to elaborate further on potential items that can be added under this domain in the revision of methodological guidelines.

- **Effect estimate**

This field presents information on the effect estimate. In the example, both the type of association (hazard ratio), the effect estimate itself, the confidence interval and p-value are shown. It should be made clear in the table that p-value are presented regardless of whether the analyses was planned a priori, pending future guideline development (see prerequisites 4.2, item 7).

[3] https://eunethta.eu/wp-content/uploads/2018/03/Direct_comparators_comparisons.pdf

[4] Guyatt et al. (2011) GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epid 64:1311-16

5.3 Information for the user and disclaimer

The Task Group proposes to add further information for the user on how GRADE domains were used in the REA, to avoid any misconception. Further, it proposed to add a disclaimer in which it is essential that it is explicitly stated that the provided information regarding the certainty of the evidence can be used, but does not need to be used (should not interfere with the conclusions on the national level). It is meant as an offer of information, but there is no requirement to use all the information. It can be supplemented with information on the used methodology and reference to the partial use of GRADE.

6 Impact of the proposal on other EUnetHTA documents

Once the Executive Board has endorsed the proposal, the Task Group suggests the following:

- A separate methodological guideline on the partial use of GRADE, with the suggested presentation, should be prepared. Authoring teams who do not have previous knowledge and experience of GRADE, should have the possibility for further assistance.
- The partial use of GRADE (as a step in the assessment phase) should be included in relevant SOPs. Creation of an own, separate SOP on GRADE could also be considered.
- Evidence tables with instructions should be included in the assessment report templates.