ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
1	ISPOR	0	overall	Overall	Should the document reflect the evolution of treatment options towards transformative and curative? Lots of the points being discussed will not apply, e.g., curative treatment options where a treatment might be one-time treatment course with limited clinical evidence and small sample study designs. All these impact the technical and clinical uncertainty of economic evalutions. Should this document make a move towards commenting on this? Not necessarily how it should be done if the team is not ready, but surely, what changes when this context is under consideration	Major	The guidelines discuss the various elements of economic evaluations (such as perspective, time horizon, and uncertainty). We make no distinction according to whether the intervention is preventive, curative, palliative, rehabilitative, or something else. We also discuss points for consideration in the event of uncertainty. However, this is not broken down according to the many different situations in which this can occur. For example, we make no value judgments about the context in which limited evidence is acceptable or not. In the summary we added some text to note that not all potentially relevant issues are tackled in this document: "*In this report, we examine points for consideration when performing/assessing an economic evaluation. Not all listed points will apply to a single evaluation. Depending on the subject, certain remarks will have more importance and others will not be relevant at all. We don't tell readers when to rely or ignore the conclusions of a study. How assessors deal with a possible identified problem is another issue and very context-specific. It is also not possible to provide a 'one-size-fit-all solution' and we cannot cover everything. Potentially relevant issues might not be tackled in this document. Technology is very broad, not restricted to pharmaceuticals. This document is a general guidance document, not specifically for one type of intervention. Nevertheless, the guidance document might help assessors decide which elements to focus on to be able to judge which evaluations are reliable or where you might ask for specific adjustments to be made. In the selection of issues and provided examples, the authors have tried to find a balance between not being too basic without becoming too technical. In doing so we hope that the document will be of practical use.* "
2	EFPIA	0			General comment: due to the different financing of health care systems accross Europe, different co-payments, different market structures and national reimbursement policies, it is doubtable that a common methodological approach for health economic evaluations is to this end puprposeful and appropriate at all.	Major	This guidance document is not about a common methodology. The national guidelines for economic evaluations remain valid. The guidance document offers support to all persons who want to review an economic evaluation critically. It is, therefore, a misconception that this report is about a common methodology for conducting economic evaluations. In the summary of the report, we have added the following sentence in the second paragraph: "*They are not used to replace or overrule national guidelines for economic evaluations.* " We also explicitly state what the intention is: "The examples are used from an educational perspective to support those who want to perform or assess an economic evaluation As mentioned by one of the reviewers after reading the first draft of this document: "*what can be learned in thousands of hours of training, reading and working is disclosed in a very attractive manner.… The idea of inserting examples from the real-world is outstanding.*"
3	EFPIA	0			Economic evaluation should not be in scope of the European cooperation on HTA (although it is acceptable that some methodological discussion/exchange of information takes place).	Major	One of the goals of EUnetHTA is also to support the reuse of national assessments. The reuse of existing economic evaluations can also be supported by indicating several essential points for consideration when critically evaluating existing evaluations. We try to do this with this report. Thus, after such a critical evaluation, it is possible to see, for example, whether the results of the study can be adopted/rejected, or if, for example, an update of the analysis is possible/necessary. We state the following: "*Whatever your background, we hope you as a reader also interpret this guideline not as a criticism of flawed approaches to economic modelling, but rather as a supportive tool for a better understanding, appropriate critical assessment and (re)use of economic evaluations.* "
4	EFPIA	0			The guideline focuses on issues beyond economic assessment and takes a very narrow perspective of economic evaluation (primarily cost utility analysis). The points for consideration and examples are helpful and perhaps this is where the real value of the guideline lies. Perhaps the authors could consider this approach to deliver a more accessible and practical guideline	Major	The scope is not limited to CUA. The part around QoL typically applies to this, but the other parts also often apply to the other types of economic evaluations. In the guidance document, we include a selection of examples in the various boxes in addition to the general description of points for consideration. In this way, we try to make the report accessible to everyone. The relevance of the points for consideration indeed depends on the type of economic evaluation people assess. For example, the part about QoL is not relevant for studies in which QoL plays no role. We mention the following in the text: "*Not all listed points will apply to a single evaluation. Depending on the subject, certain remarks will have more importance and others will not be relevant at all.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
5	EFPIA	0			Most of the guidance takes the perspective of cost effectiveness analysis primarily using cost utility. There is limited consideration of other types of economic evaluation (cost-benefit, cost-effectiveness (using life years gained), budget impact, etc). Another key observation is that much of the guideline addresses issues outside of economic evaluations that would be better addressed through reference to other guidelines on HTA, especially those focused on systematic reviews and meta-analysis. Rather than replicate the content in this guideline, a section should be written that indicates that a good economic assessment is dependent on the quality of inputs/ assumptions and that these should ideally be based on a high-quality systematic review an associated meta-analysis to provide comprehensive, unbiased data. This has several benefits especially in reducing the size of the guideline, improving focus on only economic considerations and ensuring that there is a single source reference for systematic reviews and metanalysis.	Major	The report reflects on economic evaluations in the sense of the Drummond definition: "*full economic evaluation is the comparative analysis of alternative courses of action in terms of both costs (resource use) and consequences (outcomes, effects)* ". Budget impact falls outside this definition. The comments also apply to CEA using life-years gained and other types of economic evaluations. Of course, in cases were QoL is not of importance, the part about QoL/QALYs does not apply. If there is already a EUnetHTA directive, then we refer to it at the end of each section ("*extra information* "). However, for the part on efficacy/effectiveness and safety, we have opted to add several boxes with critical pitfalls. The inclusion of examples is not a duplication of work since none of the other guidelines works with these boxes. The authors discussed this after an initial draft of the report. They preferred to make a combination of 1) referring to other guidelines where possible and 2) nevertheless adding several boxes with examples from the clinical part that have clear implications for the assessment of the economic evaluation. These examples are not exhaustive. Nevertheless, it emphasizes the importance of a critical evaluation of the clinical part when assessing an economic evaluation. In the introduction to the section on efficacy/effectiveness and safety, we write the following: "*The reliability and applicability of the results of an economic evaluation depend in the first place on the applied treatment effect and impact of adverse events. In an HTA report, safety and efficacy/effectiveness are evaluated first and provide input for the economic evaluation. The critical assessment of these elements is thus of utmost importance. Another EUnetHTA guideline elaborates on this. We refer the reader to this guideline (under construction) for further details. Nevertheless, in line with the other parts of this guideline, we refer to a selection of recommendations mentioned in other EUnetHTA guidelines and combine this with a selection of examples presented in boxes.* "
6	EFPIA	0			There is a second guideline in development on “effect and safety” – there is a significant risk for overlap between the two guidelines (as much of what is outlined here builds upon the assessment of relative effectiveness and safety). The documents spend some time discussing issues with various types of biases. There might be value to consider some “best practice” examples to balance the document.	Minor	To avoid overlap, we explicitly refer to other guidelines. We limited the part on efficacy/effectiveness and safety to several major issues. These issues are of course also important when assessing economic evaluations. In this way, we also prevent these interesting examples from being lost as the other guidelines do not work with such boxes. The title of the guidance document is "*Practical considerations when critically assessing economic evaluations* ". Writing best practice to perform economic evaluations does not fit within this framework. The focus in this document is on critically evaluating existing economic evaluations and not on conducting economic evaluations (although the points for consideration can, of course, be used to avoid specific errors when performing an economic evaluation).
7	ISPOR	0			The guideline provides a useful addition to the existing guidelines, specifically some of the examples introduced in the document. It is a rather lengthy document which some people may be put off by. Would it be possible to develop a highly summarised document alongside?	Minor	At the beginning of the text we have added a summary. See part "*Summary and table with main points for consideration* ".

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
8	ISPOR	0			Our overall opinion is, that the document is too long to be referred to as a guideline, it is also focused on issues beyond economic assessment and takes a very narrow perspective of economic evaluation (primarily cost utility analysis). There was nothing new in the guideline from what has been published extensively elsewhere. The points for consideration and examples are helpful and perhaps this is where the real value of the guideline lies. However, the examples chosen address issues that are fairly self-evident, it is the grey areas, like rare diseases or small populations or instances of limited data, that are more difficult to apply the basic principles of health economic evaluations; and, that we think readers would benefit from more insight. Perhaps the authors could consider removing much of the formal text, referencing instead other established sources, and then focus the guideline more towards a number of case studies / worked examples that would deliver a more accessible and practical guideline.	major	The added value indeed lies in the examples. We would not dare to state that these issues are self-evident as we are confronted with this when critically reviewing existing evaluations (e.g. in HTA reports or within reimbursement requests). It is not the intention of this report to make a guideline specifically for rare diseases or small populations. In each section, to avoid overlap, we also refer to existing EUnetHTA guidelines (and a limited selection of other guidelines). We keep the guidance document accessible to a very diverse audience by making a split between the general points for consideration and the boxes with examples. Someone who already has much experience with economic evaluations will immediately recognize many elements (and in other words, does not need examples to understand the issues). Someone with less experience will hopefully better understand the issues when reading the examples. The wording 'guideline' might also be confusing as this document is not really a guideline. Therefore we also changed the wording 'guideline' into 'guidance document'. For example, in the summary we mention the following: "*Whatever your background, we hope you as a reader also interpret this guidance not as a criticism of flawed approaches to economic modelling, but rather as a supportive tool for a better understanding, appropriate critical assessment and (re)use of economic evaluations.* "
9	ISPOR	0			Another overarching comment is that the language used in the document, as written, feels that the authors have assumed that the reader will have a relatively high fluency in health outcomes / health economics. Given that this is intend as a practical guide for those less technically gifted, is the document focused at the right level to be a practical guide?	major	We attempted to make the text as accessible as possible. For someone who has less experience with economic evaluations, we added a glossary (annex 4). We also refer to several references for persons who have no/less experience with economic evaluations in section 1.3: "For information on the terminology, key principles and approaches of modelling techniques, we refer to the following books (from a wide range of possible alternatives): • Methods for the economic evaluation of health care programmes. 4th ed: Oxford University Press 2015.[7] • Decision Modelling for Health Economic Evaluation: Oxford University Press August 2006.[8] • Evidence-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice. 2nd ed: Blackwell 2010.[16]"
10	ISPOR	0			The Guideline is for “economic evaluations”, however most of the guidance takes the perspective of cost effectiveness analysis primarily using cost utility. There is limited consideration of other types of economic evaluation (cost-benefit, cost-effectiveness (using life years gained), budget impact, etc). Whether intentioned or not, the guideline also implies that economic evaluations are definitive, more consideration of handling uncertainty and how this should be addressed in decision making would be valuable within a practical guideline.	major	first part of the remark: see ID5 The second part of the remark: it is outside the scope of this report to indicate how decision making should deal with uncertainty. The report does indicate what important points for consideration are when modelling uncertainty (i.e. what are possible mistakes that can be made when performing an economic evaluation).

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
11	ISPOR	0			Another key observation is that much of the guideline addresses issues outside of economic evaluations that would be better addressed through reference to other guidelines on HTA, especially those focused on systematic reviews and meta-analysis. Rather than replicate the content in this guideline, a section should be written that indicates that a good economic assessment is dependent on the quality of inputs/ assumptions and that these should ideally be based on a high quality systematic review an associated meta-analysis to provide comprehensive, unbiased data. This has several benefits especially in reducing the size of the guideline, improving focus on only economic considerations and ensuring that there is a single source reference for systematic reviews, metanalysis, etc. To further illustrate this observation, I have provided more specific comments on Sections 3.1-3.4 below. (see points that begin with ** below)	major	This comment refers to parts on efficacy/effectiveness and safety (3.1), comparator (3.2), subgroup analysis (3.3) and baseline risk (3.4). In these parts, we do not extensively repeat what these other guidelines state. Usually, only a few quotes are repeated to frame the part. We present several examples in boxes, which is not done in any of those other guidelines. There is, therefore, no overlap or duplication in the examples that we provide. As the clinical input in an economic evaluation is of primary importance, the authors have also preferred to retain these parts. As indicated by the reviewer, we also emphasize that the quality of an economic evaluation depends on the applied inputs/assumptions. The importance of unbiased information is also emphasized (see the examples in "*Box 1: The necessity of having all clinical evidence available to be able to make a proper assessment of the treatment effect* " and "*Box 2: The major problem of publication bias* " . At the end of each section, we refer to other guidelines for further information and to avoid duplication. For example, at the end of section 3.1, we refer to the following guideline: "*Extra information* • *Critical assessment of clinical evaluations. Methodological guideline. Diemen: EUnetHTA; In preparation.* "
12	ISPOR	0			The work is impressive. To me, it reseambles a text book. 155 pages, however, is quite a lot for a guideline. In general, the text would benefit from a more compact format.		The choice was made to make the text accessible and understandable for experts with different levels of experience by using examples. The actual text with points for consideration (part 3) comprises around 90 pages. A large part of these pages contains boxes with examples. After an initial internal review, several colleagues from other HTA institutes indicated that these had great added value for many reviewers and made the general comments very clear. We have opted to retain these examples and provide a summary at the beginning of the report (p8-11 "*Summary and table with main points for consideration* ").
13	ISPOR	0	general		The content has a coherent and consistent structure and clears all necessary points in the field of economic evaluations of new medicines.	major	Thank you for this comment.
14	ISPOR	0	general		List of tables facilitates rapid search of the necessary information.	major	Thank you for this comment.
15	ISPOR	0	general		A high number of examples throughout the text. We regard cases (real-world examples presented in boxes) a very helpful instrument to facilitate understanding of the key messages.	major	Thank you for this comment.
16	ISPOR	0	general		Systematic approach with clearly identified steps. The guideline presents global results and designs evaluation approach enhancements in detailing areas (e.g. transferability of economic results, economic methods and recommendations).	major	Thank you for this comment.
17	ISPOR	0	general		The references to other essential EUnetHTA, CHEERS, Drummond etc. guidelines allow navigating the way through the best practices in the field, which is also adding to the supportive role of this document.	major	Thank you for this comment.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
18	ISPOR	0	general		Detailed description of the results interpretation methods in economic evaluations (ICER threshold, in particular), which is especially important in the conditions of abcense of the formal ICER threshold in the country.	major	Thank you for this comment.
19	ISPOR	0	general		The guideline provides isssues of transferability for economic results	major	Thank you for this comment.
20	ISPOR	0	see 3.2		Not only C, but relevance of whole PICO should be discussed	minor	There is indeed a chapter entitled "comparator" and no separate title for the other elements of the PICO. However, this does not mean that we do not cover these elements in this report. For example, we mention points for consideration about the population in the section "3.4 baseline risk". A potential problem related to the baseline risk in economic evaluations is "• *Not being aware of the possible differences in the baseline risk for certain events in a specific population (e.g. selected population in an RCT) versus the population in the economic evaluation and/or the general population to which the decisions of policy makers apply. Not adjusting for such differences in baseline risk might have a large impact on the (modelled) absolute treatment effect and the ICER calculations based on it.* " To clarify that this part relates to the population, we have changed the title of this section to "3.4 baseline risk of the target population". Information about the outcome is mentioned in various other parts, for example: "*3.6 quality of life* "; "*3.7 Intermediate / surrogate versus final endpoints* ", "*3.8 Time horizon & extrapolation* ". We believe that this is clearer than trying to place these elements under the title "outcome". We identified no critical issues that are purely related to the intervention. That is why we did not write a separate chapter for this element.
21	ISPOR	0	see 3.2		In practice it is imortant to compare the new method to existing alternatives which are in current use. In Norway we recommend as comparator the therapy or method which will be displaced by the new method.	minor	It is not our intention to list all national guidelines individually. We do refer to the EUnetHTA guideline that made an overview of these national guidelines. We also note that authors, first of all, must respect national guidelines. In section 3.2, we write the following: "The first recommendation of the EUnetHTA guideline on criteria for the choice of the most appropriate comparator(s) states that: "• *Under ideal circumstances the comparator for a REA applicable across European countries should be the reference treatment according to up to date high-quality clinical practice guidelines at European or international level with good quality evidence on the efficacy and safety profile from published scientific literature, and with an EU marketing authorisation or another form of recognised regulatory approval for the respective indication and line of treatment."[12]* *However, up-to-date evidence-based clinical practice guidelines are not always available (e.g. for rare diseases) and not all interventions recommended in clinical practice guidelines are necessarily reimbursed. Researchers should thus look further and also consider current standard practice or the reimbursed alternatives, which in some cases might be different from the optimal care described in practice guidelines. Of course, in first instances, national guidelines on the choice of the comparator should be respected.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
22	PHARMERIT	0			Recommendations are missing about the level of details required for the (base case) results in order for the assessor to get a better understanding of the cost-effectiveness mechanism of the new intervention. How is the ICER build up? Where are the differences in costs and QALYs coming from? E.g. a breakdown of costs to better understand in which domain of healthcare spending the intervention might introduce savings or higher expenditures. And what's the magnitude relative to the total costs. How do costs progress over time (and what is the impact of the new intervention on that). How do utilities progress over time? What is the relative contribution of adverse event disutilities in the total QALY (difference). What is the relative contribution of survival difference on the total incremental QALYs? Etc. And can those differences be linked to the new intervention's profile and corresponding input parameters?	Major	In section 1.3, we refer to other guidelines regarding transparent reporting: "*The methodology, input, assumptions and results should be published transparently to allow a critical assessment. We refer to the CHEERS and Drummond guidelines for standards on transparent reporting, as well as to a selection of critical assessment checklists: ...* " We fully agree that transparent reporting of the input variables and the assumptions made is necessary to make a critical evaluation possible. We have added this in the introduction to the text: "*Reporting guidelines are very helpful for both researchers writing down the study results of their economic evaluation, as well as assessors identifying the relevant elements when reading such studies. Transparent reporting of the input variables and the assumptions made is necessary to enable a critical evaluation. Nevertheless, reporting guidelines do not say anything about the reliability or relevance of the results for a policymaker in a specific context. Critical assessment is the necessary next step.* "
23	EURORDIS	0	3.9		This introduction to the discount rate is insufficient in a document that should stand alone. iT is understood that national authorities will have their own rules and these must be adhered to but some guidance over understanding this is overdue. Simply stating that a higher discount rate results in a bigger impact on a calculation is just stating the obvious and offers no real guidance at all. This is, of course, not the place for a major treatise on the topic. However, EUnetHTA ought to have and to articulate some general views on the main factors to be considered when determining which discount rate to apply.	major	EUnetHTA refers to the national guidelines. In this report, we also explicitly mention to follow these national guidelines. It is not the intention to try to reach a consensus on the applied discount rate or to set up a discussion on this topic. The first point of consideration mentions the following: "• *The discount rates stated in the national guidelines for economic evaluations should be applied.* " We added the term "discount rate" in our glossary.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
24	EURORDIS	0	3.10		Together with healthcare costs and social welfare costs (direct and indirect), patients' costs should be considered as well. That includes more than item as simple ''transportation'', as patients pay for more than 60% of the costs generated by their condition (especially in rare diseases, family members must radically change thier lives to supprot people living with a rare disease (PLWRD)).	major	This report does not indicate the perspective to be taken. Here too, researchers and assessors should follow national guidelines. It is not up to EUnetHTA to impose such guidelines on national authorities. We do, however, briefly state what most guidelines say: "*Based on an overview of these guidelines, the EUnetHTA guideline on methods for health economic evaluations recommends the following:[1]* *• "Economic evaluations should at minimum be conducted from a health care perspective. However, several countries require a societal perspective. Presenting the use of resources as related to other sectors of society may increase the usefulness of the analysis to more EUnetHTA partners. Regardless of perspective taken, it is recommended that the use of resources is presented in as detailed a manner as possible. For example, if a societal perspective is used, indirect costs should be presented separately."* " This also does not mean that costs for patients are not included. For example, in one of the points for consideration we mention the following: "*• The perspective defines the perimeter of consequences of health interventions to consider. It should be transparent which outcomes and costs are studied (e.g. only within or also outside the health care sector), whose outcomes (e.g. only for the patient or also for the caregiver or society) and which costs are studied (e.g. are only costs for the government included or also patient's co-payments and other costs). For example, in the Belgian guidelines for economic evaluations, "the identification, measurement and valuation of costs should be consistent with the perspective of the Belgian health care payers."[92] In this guideline, 'health care payers' refers to both the patients, the federal government and the communities.* "
25	EURORDIS	0	3.11		The calculation and assessment of incremental costs can include several types of costs, for instance: intervention costs that are directly related to the studied interventions, cost savings that arise as a result of the effectiveness of the interventions, cost increases due to adverse effects related to the studied intervention, cost increases or savings from follow-up treatments	major	We changed this sentence as follows: "The incremental costs ~~contain~~ can include several types of costs, for instance: ... "
26	ISPOR	9	Summary	Table - last bullet point	"information" should be clarified. There is lots of literature and guidance as how this should be collected and how this should be collected, so a value reference to "information" is not advancing the status quo. There are many initiatives to capture patient experience, patient satisfaction and other patient reproted outcomes across different stakeholders including FDA, hence there is a need for this bullet point to come forward as an important element for economic evaluations to evolve, otherwise, it reads like this document is a summary of different considerations and points of attention yet with limited effort to recommend an improvement and evolution of the science behind it	Minor	Sections 3.6 and 3.7 provide further information based on existing guidelines. We refer to the existing EUnetHTA guidelines that contain further information. "*Extra information* *• Endpoints used for Relative Effectiveness Assessment: health-related quality of life and utility measures. Methodological Guideline: EUnetHTA; 2015.[10]* " "*Extra information* *• Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints. Methodological Guideline: EUnetHTA; 2015.[13]* *• Endpoints used in Relative Effectiveness Assessment: Surrogate Endpoints. Methodological Guideline; 2015.[15]* " This is also further explained in the points for consideration. For example, we mention the following in one of these points: "*This is in line with the EUnetHTA recommendations stating that "mapping of disease-specific or generic instruments to preference-based instruments to obtain utility values is generally not recommended for REA. Authorities should encourage researchers to always include a preference-based instrument in their clinical trial protocol in order to avoid the need for mapping."[10]* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
27	ISPOR	9	Summary	Table - last bullet point	The table is a mix of recommendation and observations. Not sure what the aim of the table is if no recommendation are to be given in this summary. It reads like no recommendation is given for some aspects, hence it might prevent scientific audience to engage	Major	This table does not contain any 'recommendations'. It is an overview of possible points for attention. We mention the following above the table: "*The following table provides an overview of the elements that are discussed in this report together with a selection of points for consideration. For more information we refer to the respective parts of the full report.* "
28	EURORDIS	9	Summary and table with main points 1 for consideration	ble in summa	Efficacy and safety: Was the assessment of the efficacy/effectiveness carried out according to current standards? For efficacy and safety, the quality of evidence as rated by current standards are already evaluated by regulatory agencies for pharma, and notified bodies for medical devices. Maybe other technologies' efficacy and safety are not assessed when they come to HTA. But when regulators already evaluated the quality of all studies submitted for the uathorisation of the product, according to international standards (e.g. guidelines from International Conference on Harmonisation), why to repeat this in an HTA report? Why to re-assess quality of data according to current standards?	major	This report is about HTA, to support, for example, reimbursement decisions. There are many essential differences in the evaluation of a drug to obtain registration or a medical device to obtain a CE label versus the evaluation in a reimbursement request. For example, for registration, experts assess the benefit-risk ratio in which comparators or endpoints may be different (or have a different level of importance) in comparison with a reimbursement file. To obtain a CE label, the legislation mentions proving safety and performance, which is not the same as demonstrating an added value compared to existing alternatives. The examples we provide in the boxes also show that it is important to perform a good critical evaluation of the clinical part (safety & efficacy/effectiveness) when assessing an economic evaluation, even if there is already a registration or a CE label.
29	PHARMERIT	10	-	-	Referenced section "discount rate" needs to be updated	Minor	The cross-reference to part 3.9 is added.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
30	ISPOR	10	Summary	Table - Model verification and validation	Model validation is not properly described. The model validation is a broad exercise and is not limited to whether assumptions reflect reality. This is validation of a case/ assumption/ strategy, not the model's	Major	Indeed, we did not include all possible nuances in the summary. We cover such nuances in section 3.13. We also mention in the text that different typologies are used in the literature. "*To assess how good a model is, we must ascertain whether the model implements the assumptions correctly (model verification) and whether the assumptions which have been made are reasonable and reflect reality (model validation).* *• Verification is concerned with the technical accuracy of the model and should identify "programming errors, data entry errors, and logical inconsistencies in the model specification."[177]* *• Validation is concerned with the structure, content and predictive accuracy of the model.* *The HTA Core model (2013) specifies "to be able to evaluate how the results of a model should be used, a model user benefits from knowing how well the model predicts the outcome(s) of interest. To be able to do this, the model needs to be transparently reported and validated." "Validation relates to methods to evaluate the extent of a model's accuracy in making relevant predictions or in abstracting from a complex reality". "In cases where validation is possible, e.g., using a relevant data set, it is recommended".* *The following verification and validation exercises should be explored:* *• Face validity: Does a model structure, its assumptions, input parameter values and distribution and output values and conclusions, make sense and can be explained at an intuitive level?* *• Internal validity (technical verification): Has the model been implemented correctly?* *• Cross model validation: Does the model achieve similar results with other models that were independently developed, but aimed at estimating the same outcomes?* *• External validation: How can we compare the outputs of the model with actual outputs provided by external sources (not used in the model)? If a source for future events is available, Eddy et al.[180] define "predictive validation".* " The following footnote refers to alternative typologies: "*Other typologies have been proposed in the literature.* *- Eddy et al; (1985)[178]: [1] First-order validation requires expert concurrence; [2] Second-order validation compares the model predictions with data used to estimate the model parameters; [3] Third-order validation compares the model prediction with "other" observed data, i.e. data not used in the model construction; [4] Fourth-order validation compares pre-implementation model predictions with observed events post-implementation.* *- Vemer and al. (2016)[179]: [1] conceptual validation, [2] data validation, [3] Computerized model validation, [4] Operational validation.* "
31	PHARMERIT	12	1.1	20,25	Consistently use either "policymaker" or "policy maker"	Linguistic	This is corrected. We now consistently use 'policy maker'
32	ISPOR	13	2	10 - 12	The quality of evidence is a key factor and parameter that is being considered already in economic evaluations and is key driver of the validity and feasiblity of an evaluation. This should be added in the elements	Major	These elements are indeed essential. They are also further elaborated in another guideline. We refer to this guideline to avoid overlap. Since these are very important elements in the assessment of economic evaluations, we have nevertheless chosen to include a selection of examples in boxes (in section 3.1). We mention the following in the report: "*Two very important elements (treatment effect and safety) will be elaborated on in a separate guideline (Critical assessment of clinical evaluations – under construction) and are thus not elaborated on in detail in this report. For these elements we will only include a limited selection of examples of points for consideration and refer to the more detailed guidelines (under construction).* " In the introduction of part 3.1, the following is mentioned: "*The reliability and applicability of the results of an economic evaluation depends in the first place on the applied treatment effect and impact of adverse events. In an HTA report, safety and efficacy/effectiveness are evaluated first and provide input for the economic evaluation. The critical assessment of these elements is thus of utmost importance. Another EUnetHTA guideline elaborates on this. We refer the reader to this guideline (under construction) for further details. Nevertheless, in line with the other parts of this guideline, we refer to a selection of recommendations mentioned in other EUnetHTA guidelines and combine this with a selection of examples presented in boxes.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
33	PHARMERIT	16	-	-	Referenced section "discount rate" needs to be updated	Minor	The cross-reference to part 3.9 is added.
34	ISPOR	18	3.1.1	18-19	The section indicates efficacy/ effectiveness, yet only efficacy is being covered. It should be clear that efficacy and effectiveness are not the same and effectiveness should be treated differently than efficacy when available, mentioning all the limitations of using effectiveness in economic evaluations	Minor	Efficacy and effectiveness are indeed not the same. We have added the definition of both terms in our glossary: "*Effectiveness: The extent to which a specific intervention, when used under ordinary circumstances, does what it is intended to do. Clinical trials that assess effectiveness are sometimes called pragmatic or management trials. (Cochrane glossary)* *Efficacy: The extent to which an intervention produces a beneficial result under ideal conditions. Clinical trials that assess efficacy are sometimes called explanatory trials and are restricted to participants who fully co-operate. (Cochrane glossary)*" It is not correct to state that there are anyhow restrictions when researchers use effectiveness information. If such information comes from a well-performed pragmatic RCT, then this information can be very reliable and applicable. The reviewer probably wants to refer to the problem with observational data (which is sometimes incorrectly labelled as effectiveness information). The problem with such data is addressed in "Box 3: The (questionable) reliability of evidence on the relative treatment effect from non-randomized studies". If the reviewer refers to the difference in patient characteristics between an RCT and the actual population, then this is discussed, for example, in section 3.4. "*Potential problems related to the baseline risk in economic evaluations are: • Not being aware of the possible differences in the baseline risk for certain events in a specific population (e.g. selected population in an RCT) versus the population in the economic evaluation and/or the general population to which the decisions of policy makers apply. Not adjusting for such differences in baseline risk might have a large impact on the (modelled) absolute treatment effect and the ICER calculations based on it. ...* "
35	EURORDIS	18	3.1.1	33_38	The introduction to Box 3 example and the select quote of EUnetHTA guidelines, do not take into consideration conditions or situations where a RCTs cannot be run (e.g. for ethical reasons), or the possibility to observe a clear benefit via a NRS (e.g. in overall survival, on in quality of life improvement), depending on factor as population, severity of the condition and feasibility, and benefit to be measured.	major	This is indeed the case. We have added the following nuance in a footnote: "*This does not exclude that in some cases, it is impossible to perform RCTs, for example, for ethical reasons or because a clear benefit was observed through non-randomized studies. A frequently quoted non-medical example notes that no randomized studies have been performed for parachutes.(REF: Smith, BMJ, 2003) On the other hand, it is also important to point out that several medical examples do not stand the comparison with the parachute example. A study evaluated claims that a medical practice is akin to a parachute. The authors conclude that "Although we found that the parachute analogy is seldom used to describe a medical practice, when it is used it is often inappropriate, incorrect or misused" (REF: Hayes, CMAJ, 2018). But also in this paper, the authors nuance by mentioning that their paper "does not imply that RCTs are always feasible, possible, necessary or ethical." (REF: Hayes, CMAJ, 2018)* "
36	EFPIA	18	3.1		The content of chapter 3.1 Efficacy/effectiveness and safety relates to other EUNetHTA guidelines and should therefore be dealt with in these respective guidelines to avoid confusing double information and keep guidance consistent	Major	zie ID11

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
37	EFPIA	18	3.1		This section is not focused on consideration of efficacy/effectiveness. It is actually addressing the need for a comprehensive, systematic literature review that gives precedence to high quality studies. This is important, but is not specifically focused on issues related to how efficacy is measured across different studies, reliability of endpoints, measurement of adverse events, how these are analysed (meta-analysis, mixed treatment comparisons, etc) and how these are included in an economic analysis, etc. These topics are absent from these sections but would seem more pertinent to a guideline on economic evaluation to ensure you are including the right data. As written, the section would fit better with a guideline on systematic literature reviews. If the latter was the intention, then much of this section could be removed and the guideline simply indicate that all data for an economic analysis should come from a comprehensive systematic literature review and then reference other guidelines on this topic.	Major	It is not the intention of this document to indicate how to measure efficacy, how to perform a good meta-analysis, etc. These are all elements outside the scope of this report. As indicated above, we refer to other EUnetHTA guidelines where possible. We have included several boxes with examples of problems concerning efficacy/effectiveness and safety. There is no duplication because such examples were not worked out in other guidelines. Concerning reliability of endpoints: part 3.7 is about 'Intermediate/surrogate versus final endpoints' we have also added the reference to the systematic review guideline: "E*xtra information* • *Critical assessment of clinical evaluations. Methodological guideline. Diemen: EUnetHTA; In preparation.* • *Institute for Quality and Efficiency in Health Care (IQWiG). Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness Methodological Guideline; 2017.[22]* "
38	EFPIA	18	3.1		Issues relative to safety that should also be covered here are how to cost these events and ensure that they are fully captured in the model, especially if resource use isn't collected in the trials. Important that rare, but potentially costly, events are not overlooked. The frequency of these events might be difficult to determine from RCTs, so there is a need to also include real world data.	Major	The objective of this report is not to say how and where researchers should collect information or how they should value costs. The main intention is to provide points for consideration when evaluating an existing economic evaluation. In part "3.11 (Context-specific) costs" we provide a non-limitative list with points for consideration. We added the following to this list: "• *It is important to verify that all important incremental elements have been sufficiently taken into account. For example, have severe adverse effects, which may not be common but can cause high costs (and impact QoL), been satisfactorily included (see part 3.1.2).* " In part 3.1.2 we also mention the following where we refer to the 'real-world data' as the reviewer correctly states: "• *The data on adverse effects may come from different sources with different risk of bias. Ideally the safety profile of the technology should be described against the comparator and those clinically significant differences in adverse effects between the technology and the comparator should be considered in the analysis. However, observational studies may be particularly useful for long-term or infrequent adverse effects. On the other hand, this may not provide an unbiased source of comparative safety information.* "
39	EFPIA	18	3.1		Treatment effect: We would like to have seen more guidelance on non-randomized trials with single arms, no placebo and small numbers. This is particularly topical given the emerging gene therapies and regenerative medicine where the numbers in the trials may be less that 15-20 patients and trial duration is short.	Minor	This is out of scope of this report.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
40	ISPOR	18	3.1		Treatment effect: We would like to have seen more guidance on non-randomized trials with single arms, no placebo and small numbers. This is particularly topical given the emerging gene therapies and regenerative medicine where the numbers in the trials may be less that 15-20 patients and trial duration is short.	Minor	This is out of scope of this report.
41	ISPOR	18	3.1		** This section is not focused on consideration of efficacy/effectiveness. It is actually addressing the need for a comprehensive, systematic literature review that gives precedence to high quality studies. This is important, but is not specifically focused on issues related to how efficacy is measured across different studies, reliability of endpoints, measurement of adverse events, how these are analysed (meta-analysis, mixed treatment comparisons, etc) and how these are included in an economic analysis, etc. These topics are absent from this section but would seem more pertinent to a guideline on economic evaluation to ensure the right data are included. As written, the section would fit better with a guideline on systematic literature reviews. If the latter was the intention, then much of this section could be removed and the guideline simply indicate that all data for an economic analysis should come from a comprehensive systematic literature review and then reference other guidelines on this topic. Safety considerations are better articulated in this section, but could be further developed. Issues relative to safety that should also be covered here are how to cost these events and ensure that they are fully captured in the model, especially if resource use isn't collected in the trials. Important that rare, but potentially costly, events are not overlooked. The frequency of these events might be difficult to determine from RCTs, so there is a need to also include real world data.	major	See ID37 and ID38
42	ISPOR	18	3.1.1		It seems, that estimating outcomes and cost effectiveness using a single-arm clinical trial is not covered in this section.	major	The report includes a point of consideration that is related to this and is elaborated in "Box 3: The (questionable) reliability of evidence on the relative treatment effect from non-randomized studies"

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
43	EURORDIS	18	3.1		Same comment: "In an HTA report, safety and efficacy are evaluated first and provide input for the economic evaluation". No reference is made to the available assessment reports of efficacy and safety/toxicity by regulatory bodies for pharmaceuticals, and notified bodies for medical devices. Depending when the HTA is made, additional information might be available compared to the regulatory report, but for products authorised via the centralised procedure in the EU, all clinical study reports can be available on request , published or not (Access to document policy 043 at EMA)	major	See ID28
44	EURORDIS	19	3.1.1	1_4	The introduction to Box 3 example and the select quote of EUnetHTA guidelines, do not take into consideration conditions or situations where a RCTs cannot be run (e.g. for ethical reasons), or the possibility to observe a clear benefit via a NRS (e.g. in overall survival, on in quality of life improvement), depending on factor as population, severity of the condition and feasibility, and benefit to be measured.	major	See ID35
45	EFPIA	19		42 & 18-22	The guideline does not discuss pros and cons of indirect evidence in a balanced way. It is true that there is always a varying amount of uncertainty but including indirect evidence on comparators that are relevant for the decision problem is very important from the decision makers' point of view. This is brought up in the section 3.2. Comparators, which emphasises the importance of including all relevat compartors. In the appraisal of a new technology in the competitive therapeutic are it also inevitable that direct evidence is insufficient and appropriate indirect evidence is needed to perform valid incrementat CE-analysis. Currently the text gives an idea that indirect evidence is "nice to have" while in reality it is, in many cases, a necessity.	Major	The focus of this report is to provide points for consideration. It does not provide pros and cons or indirect evidence. Also for this element, the report shortly refers to the existing EUnetHTA guideline on this topic and provides an example in Box 4: "*Bias in head-to-head comparisons and uncertainty linked to evidence relying on indirect comparisons instead of direct comparisons is also an important issue which should be taken into account when interpreting results of economic evaluations. In Box 4, we provide an example of bias in head-to-head comparisons were the outcomes depended on the study sponsor. We also refer to the EUnetHTA guideline on direct and indirect comparisons[11] which states that:* • *"The choice between direct and indirect comparison is context specific and dependent on the question posed as well as the different evidence available. Where sufficient good quality head-to-head studies are available, direct comparisons are preferred as the level of evidence is high. Should substantial indirect evidence be available, then it can act to validate the direct evidence. When there is limited head-to-head evidence or more than two treatments are being considered simultaneously, the use of indirect methods may be helpful."* • *"An indirect comparison should only be carried out if underlying data from comparable studies are homogeneous and consistent, otherwise the results will not be reliable."* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
46	ISPOR	19		42 & 18-22	The guideline does not discuss pros and cons of indirect evidence in a balanced way. It is true that there is always a varying amount of uncertainty but including indirect evidence on comparators that are relevant for the decision problem is very important from the decision makers' point of view. This is brought up in the section 3.2. Comparators, which emphasises the importance of including all relevant compartors. In the appraisal of a new technology in the competitive therapeutic are it also inevitable that direct evidence is insufficient and appropriate indirect evidence is needed to perform valid incrementat CE-analysis. Currently the text gives an idea that indirect evidence is "nice to have" while in reality it is, in many cases, a necessity.	Major	see ID45
47	EURORDIS	21	3.1.1	Box 2	The conclusion of Box 2 example seems to set up a correlation between the negative outcome of the assessment and the publication bias. The bias is real and well-documented, but the rationale for the negative outcome is not presented.	minor	It is a finding made in various papers. The rationale behind it is not explained. Based on expert opinion, some state that the researchers are not inclined to publish negative results. According to others, it is because it is more difficult to publish negative results. The most important thing for us is the observation of publication bias, together with the last sentence in this box: "*All efforts to identify (e.g. by searching all relevant studies in trial registries) and retrieve all evidence, inclusive results from non-published studies, should be supported to allow better unbiased estimates of the treatment effect.* "
48	EURORDIS	23,24	3.1.1	Box 3	The introduction to Box 3 example and the select quote of EUnetHTA guidelines, do not take into consideration conditions or situations where a RCTs cannot be run (e.g. for ethical reasons), or the possibility to observe a clear benefit via a NRS (e.g. in overall survival, on in quality of life improvement), depending on factor as population, severity of the condition and feasibility, and benefit to be measured.	major	see ID35 & 44

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
49	EURORDIS	27	3.1.2		About safety: the definitions of" harm" and "averse event" are useless. The correct terms should be "toxicity and risks" profile and not "just safety". Similarly, prefer "potential benefits" than benefits when discussing efficacy. For causality, use the terms "adverse event" , or "adverse reaction" when causality is possible, probable or demonstrated. The medical definition of "harm" is "Anything that impairs or adversely affects the safety of patients in clinical care, drug therapy, research investigations, or public health. Harms include adverse drug reactions, side effects of treatments, and other undesirable consequences of health care products and services." harms. (n.d.) Medical Dictionary. (2009). Retrieved October 11 2019 from https://medical-dictionary.thefreedictionary.com/harms Therefore the scope of "harm" is limited: it it limited to damage that has occurred, it does not consider possible risks. Using the term "harm" minimises the problems a technology might cause. A correct term would be "risks": it includes both damages that occur, and possible ones. For teratogenic products for example: harm is when a child is born with abnormalities. Risks are the possibility that babies are born with abnormalities if exposed during pregnancy. Risks can be extremely high (e.g thalidomide), even if harm is null (as long as all women do not take thalidomide during pregnancy, no baby is born with abnormalities). Thus, it is suggested to replace "safety" (preferred term by industry) by "toxicity" and "harm" by "risks" in	major	We acknowledge that several different definitions are used in the literature. In the text, we do indeed refer to the term that is often used in the context of 'harms & benefits', which is not limited to 'damage that has occurred'. In the text, we have sought consistency by citing the definitions used in the EUnetHTA safety report. We also refer to this report as follows: "*To see a list of terms and definitions, please refer to the EUnetHTA guideline 'Endpoints used in Relative Effectiveness Assessment: Safety'*. " This EUnetHTA guideline about safety contains three definitions on harm: "*Harms* *The totality of possible adverse consequences of an intervention or therapy; they are the direct opposite of benefits, against which they must be compared. (Source: Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004; 141(10):781-788.)* *The nature and extent of actual damage that could be caused by a drug. (Source: WHO. Glossary of terms used in Pharmacovigilance.)* *In the context of MD regulation: Physical injury or damage to the health of people, or damage to property or the environment. (Source: Guidelines on medical devices. Vigilance System. MEDDEV 2.12-1 rev. 8 2013.)* " We adopted the first definition in our report. This definition includes 'the totality of possible adverse consequences', and thus does not limit to 'damage that has occurred' and does not minimises the problems a technology might cause. We also added 'adverse effect' & 'adverse event' in the glossary referring to the terminology used in the Cochrane glossary since their definitions also explain directly what is meant by adverse drug reaction and adverse reaction. As the literature uses the terms safety & harms rather than toxicity & risks, we preferred not to replace these terms in the document systematically.
50	EFPIA	28		24	Related to previous comment, the statment that "…tend to not suffer RCT's limitations..." gives somewhat unbalenced view. While it is unarguably true, also RCT's have other advanges also as source of AE data.	Minor	the sentence refers to the limitations that were mentioned in the previous sentence. We added 'the above RCT limitations' to make this clear: "*RCTs may not be generalizable as they tend to exclude patients at higher risk of adverse effects. Their usual short-term follow-up and sample size may reduce the likelihood of appearance of adverse events. Observational studies are very useful for the observation of adverse events as they tend to not suffer from the above RCT limitations.* "
51	ISPOR	28		24	Related to previous comment, the statement that "…tend to not suffer RCT's limitations..." gives somewhat unbalanced view. While it is unarguably true, also RCT's have other advantages also as sources of AE data.	minor	see ID50
52	EFPIA	28		36	Also, should be noted that "infrequent AEs" may not have any meaningful inpact on CE-model outcomes, and ICER in particular - unless they are expected to have significant implications on QoL and/or resource use.	Major	OK, in the part on costs, the following has been added: "• *It is important to verify that all important incremental elements have been sufficiently taken into account. For example, have severe adverse effects, which may not be common but can cause high costs (and impact QoL), been satisfactorily included (see part 3.1.2).* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
53	ISPOR	28		36	Also, should be noted that "infrequent AEs" may not have any meaningful inpact on CE-model outcomes, and ICER in particular - unless they are expected to have significant implications on QoL and/or resource use.	major	see ID52
54	EFPIA	28	3.1.2.	34-36	Similar cautioness than with the efficacy/effectivenss based on the NRS/observatoinal studies (p.18, ln 37) should be added here. E.g. "Although observational studies may be particularly useful for long-term or infrequent adverse effect they may not be an unbiased source of comparative safety"	Major	OK, we added this in the text: "*… observational studies may be particularly useful for long-term or infrequent adverse effects. On the other hand, this may not provide an unbiased source of comparative safety information.* "
55	ISPOR	28	3.1.2.	34-36	Similar cautions as with the efficacy/effectiveness based on the NRS/observational studies (p.18, ln 37) should be added here. E.g. "Although observational studies may be particularly useful for long-term or infrequent adverse effect they may not be an unbiased source of comparative safety"	major	see ID54
56	EURORDIS	28	3.1.2	7_14	*" 'it is clear from the available guidance that all relevant outcomes should be included in the economic decision model and there appears to be a general if not clearly stated consensus that this includes adverse effects' but 'articles contained very little information or guidance of direct relevance to the incorporation of adverse effects in models' ".* The 4th edition of CADTH Guidelines *Guidelines for the Economic Evaluation of Health Technologies: Canada* suggest a way to include AE in a cost-utility model (https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf. See p. 36 and p. 38-39)	major	We added the part of the CADTH guideline on adverse events: "*The Canadian guidelines for the economic evaluation of health technologies provide the following clear description:(REF CADTH 2017)* • *"Researchers should be explicit about how the adverse events included in the economic evaluation were identified, and what methods were used to incorporate them. Where adverse events have a negligible impact on health effects, or no impact on costs and resources, it is often appropriate to exclude these events from the model. Where adverse events are not included, a clear justification must be provided.* • *Adverse events should be incorporated into the model by combining both the health condition and the associated adverse effects. In the case of utilities, the utility for a specific health state can then be adjusted by applying a disutility for an adverse event to allow the utility for the health state with an adverse event to be estimated.REF* • *If effects are transitory (i.e., short-term), they should be incorporated through appropriate refinement of the states or events within the model. Where data are available on the prevalence, costs, and disutility associated with each adverse event by intervention, this facilitates greater transparency. "*
57	EFPIA	30	3.2		Comparators: We strongly belive that off-label medicines must not be included as comparators in economic evaluations	Major	Off-label use is mentioned in several national guidelines as an appropriate comparator under certain circumstances. We refer to several examples in the footnote (Belgium, Ireland, UK & Poland). Similar to all other elements, we state that in the first instance, the national guidelines must be consulted and followed.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
58	EFPIA	30	3.2		Similar to section 3.1 much of this is focused on undertaking a high-quality systematic review and could be summarised with reference to the appropriate guidelines Important issues that aren't currently addressed in this section relevant to the comparator are how to determine relative effectiveness between the treatment of interest and the relevant comparators. This is the crux of the issue relevant to undertaking an economic analysis. To restate: The systematic review will identify the relevant comparators, what is important for the economic guideline is to establish how you then generate the relevant data to populate your model.	Major	The authors preferred to start for all parts with a brief reference to existing guidelines. Several points for consideration are then listed. After that, we provide two real-world examples about 'Problems related to the inappropriate ex- or inclusion of alternatives'. This is not an overlap with other guidelines. We also refer to the EUnetHTA guideline on comparators: "• *Comparators & comparisons: Criteria for the choice of the most appropriate comparator(s). Methodological Guideline: EUnetHTA; 2015* "
59	ISPOR	30	3.2		Comparators: We strongly believe that off-label medicines must not be included as comparators in economic evaluations	Major	see ID57
60	ISPOR	30	3.2		**Similar to section 3.1 much of this is focused on undertaking a high quality systematic review and could be summarised with reference to the appropriate guidelines. Important issues that aren't currently addressed in this section relevant to the comparator are how to determine relative effectiveness between the treatment of interest and the relevant comparators. This is the crux of the issue relevant to undertaking an economic analysis. To restate: The systematic review will identify the relevant comparators, what is important for the economic guideline is to establish how you then generate the relevant data to populate your model.	major	see ID58
61	EFPIA	31		12-13	It is not clear "…other new interventions…" refers also new technologies introduced and assessed within a relatively short period of time? In this case it should be noted that typically for non-established technology the cost of intervention may not be available and also other detailed data needed for comparison is likely to be limited.	Minor	This is indeed the case: everything stands and falls with the availability of reliable information. However, since we think this is rather obvious, we did not add this explicitly.
62	ISPOR	31		12-13	It is not clear "…other new interventions…" refers also to new technologies introduced and assessed within a relatively short period of time? In this case it should be noted that typically for non-established technology the cost of intervention may not be available and also other detailed data needed for comparison is likely to be limited.	minor	see ID61

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
63	ISPOR	31	3.2	18-24	Reference to treatment guidelines is not always sufficient and is not always how economic evaluations are performed. Next to the treatment guidelines (when available), clinical experts and physicians and patients' view is being considered and this should be reflected in this document as well along with the cautiousness we should be mindful of when doing that	Major	we added the following sentence in this part: "*Expert opinion or patients' view might also be helpful in identifying relevant comparators.* "
64	EURORDIS	31	3.2	1_17	Among 'Point for consideration', it would be meaningful to clarify what follows : the decision of comparing a new technology with an existing one (different from the best clinical practice) because of its cost-effectiveness ratio, should be justify further. Whenever possible, the (considered cost-effective) comparator should not be chosen only among those for the same or a similar condition, but also among those intended to bring the same benefit of the one under evaluation. On this point, it becomes crucial to discuss with patients.	major	It is not correct to state that an intervention should only be compared with other interventions that hopefully yield the same benefit. If both interventions are not cost-effective at all, then it is wrong, for example, to automatically exclude the intervention that has better cost-effectiveness (even if this intervention is inferior). We elaborate on the issue of incorrect exclusion of interventions in box 7 & 8 with two practical examples. "*For example, in a study evaluating the use of tumour necrosis factor-alpha (TNF-a) inhibitors, adalimumab and infliximab, for Crohn's disease, the authors explain why it is important not to compare only biologicals with each other. This would only be relevant "where both adalimumab and infliximab have been first justified as maintenance therapies versus standard care (SC). Where one or both maintenance therapies are not cost-effective versus SC, this comparison provides no information to decision-makers."[87]* "
65	ISPOR	32	see 3.2	4	or not previously assessed with respect to cost effectiveness	major	we added this in our text: "• *It is possible that standard of care is not cost-effective (or that the cost-effectiveness has not been assessed previously) ...* "
66	INFARMED	32	3.2.	4	It is unclear why a cost-effective alternative should be found as comparator when the standard of care is not cost-effective. This may mean to consider, as comparator, a therapy that is not part, or that is weakly representative of standard of care in the country. This is not an easy decision. Same question in box 7: was the best comparator, CRT-P, part of standard of care in Belgium?	minor	This is indeed a point for consideration. It is important to take into account more cost-effective alternatives, even if they are not standard of care, when carrying out or critically assessing an economic evaluation. The example of CRT-P and CRT-D is an excellent example since an RCT included both treatment alternatives. If some experts only want to evaluate the (more expensive) CRT-D without taking into account the (cheaper) CRT-P, then this can lead to incorrect findings and conclusions. If an alternative is not standard of care, but it appears that based on the latest evidence about safety, efficacy and cost-effectiveness it is a better alternative than the current standard of care, than conclusions and recommendations can take this into account.
67	ISPOR	32	see 3.2	13	In practice it might be challenging to assess all relevant alternative subgrpups since the recources are constrin and the timereframe is often limited	minor	It is indeed possible that the comparator can be a mix of various alternative interventions. However, in some cases, the comparator is very clearly different. Such an example is provided in the paragraph starting with "When the comparator is not the same across the target population". The example given in the last sentence provides clarification: "*e.g. in aortic stenosis, some patients are inoperable (→ optimal medical treatment as comparator), while others are operable (→ surgery as a comparator)* "
68	ISPOR	32	see 32.2	24	and sales figures	minor	We added this in our text: "These alternatives might be identified through a systematic review of the medical literature, clinical practice guidelines, expert opinion, patients experience and perspective, sales figures, etc."
69	EURORDIS	32	3.2	1_25			?

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
70	EURORDIS	32	3.2	22_25	*"Identifying potentially relevant alternatives is usually part of the clinical section of an HTA. These alternatives might be identified through a systematic review of the medical literature, clinical practice guidelines,* ***patients experience and perspective*** *, expert opinion, etc. Alternatives for which no evidence is available are difficult to evaluate in an economic evaluation. "*	major	see ID68
71	EFPIA	36	3.3		Agree that this is an important consideration of economic analysis. However, the text again is more relevant to subgroup analysis within the context of a systematic review and addresses issues relevant to clinical considerations as much as economic ones. Much of the context could be simply cross references to a guideline on systematic reviews and meta-analysis. With regard to economic consideration of subgroups I would have expected to see commentary that subgroup analysis is worth consideration if cost-effectiveness isn't proven for the whole population. I would also have thought commentary on statistical testing for interaction is an important consideration to support the validity of the defined subgroups.	Major	It is indeed the case that subgroup analysis might be an important element in the clinical assessment. However, this might also be of great importance for the economic evaluation and we make a link with this economic aspect in various points for considerations. It is not correct to state that subgroup analysis is only interesting if the intervention proves not to be cost-effectiveness for the whole population. Even if cost-effectiveness has been proven for a specific population, it may still be the case that an intervention is not cost-effective for a particular subgroup within this population. For example, if there is evidence within the target population that a subgroup has a high risk of developing a major adverse event (e.g. heart failure in patients with a relatively low left ventricular ejection fraction (LVEF)), then it may be useful to take this into account in the economic section and to split the population according to the initial LVEF status. We mention this in one of the comments (also referring to section 3.4). "• *The cost-effectiveness of an intervention will be different for subgroups if the relative treatment effect differs. However, the heterogeneity of the absolute treatment effect is also of importance. In this context, it is important to consider other factors such as sociodemographic characteristics (e.g. age, sex, social class) or clinical characteristics (e.g. baseline risk (see part 3.4) or disease severity).* "
72	ISPOR	36	3.3		**Agree that this is an important consideration of economic analysis. However, the text again is more relevant to subgroup analysis within the context of a systematic review and addresses issues relevant to clinical considerations as much as economic ones. Much of the content could be simply cross references to a guideline on systematic reviews and meta-analysis. With regard to the economic considerations of subgroups, I would have expected to see commentary that subgroup analysis is worth consideration if cost-effectiveness isn't proven for the whole population. I would also have thought commentary on statistical testing for interaction is an important consideration to support the validity of the defined subgroups.	major	see ID71

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
73	ISPOR	37	3.3	12-18	We consider the bullet point invalid. If from the clinical and patient point of view, a subgroup is performing differently and should be treated differently, decision-makers should accommodate the unmet need. We don't get to see the message that is being forward as useful to the scientific audience as the economic evaluations should inform the decision-makers about all options, benefits and disadvantages even given for a subpopulation that may not be feasible to treat, yet they should.	Major	We agree with the comment and have chosen to delete this bullet point.
74	EFPIA	38	3.4		This entire section as written is relevant to meta-analysis and clinical assessment. How does baseline risk impact an economic evaluation? Again, much of this section should be simply referenced to other guidelines. Ideally, the section would start with the bullets under points for consideration, provide the example and then reference the relevant section in an associated guideline on systematic reviews and meta-analysis.	Major	It is not correct to state that this is only relevant in the context of a systematic review. It may be perfect for a systematic review of the clinical literature to collect information on the relative and absolute treatment effects in RCTs while analysis of real-world data provides information on the baseline risk for specific problems in the real-world population treated with the selected comparator.
75	ISPOR	38	3.4		**This entire section as written is relevant to meta-analysis and clinical assessment. Where is the discussion regarding how baseline risk impacts an economic evaluation? Again much of this section should be simply referenced to other guidelines. Ideally, the section would start with the bullets under points for consideration, provide the example and then reference the relevant section in an associated guideline on systematic reviews and meta-analysis.	major	The text gives a hypothetical example of an incorrect absolute benefit: "*To illustrate this with a hypothetical simple example: if a trial shows that an intervention reduces the one-year mortality rate in a specific indication from 20% to 12%, but in reality, the one-year mortality rate in this indication with the current treatment is only 5%, then of course you cannot avoid an absolute 8% (or in other words: 8 percentage points) of deaths. Similarly with relative outcomes: if a trial shows a baseline event rate occurring in 30% of the population in the comparator arm and the intervention has a relative risk (RR) of two, then we see the event in 60% of the population in the intervention arm. If our baseline event rate in the real-world population was 55% then simple application of the same RR would give us an event in 110% of the population after introduction of the intervention, which is of course not possible. Applying incorrect or unrealistic absolute benefits in economic evaluations will of course result in non-reliable cost-effectiveness outcomes.* " We have chosen not to give an example where a researcher has made such a 'mistake' since this would not be comfortable for the author. We have added the following sentence after this paragraph: "*Applying incorrect or unrealistic absolute benefits in economic evaluations will of course result in non-reliable cost-effectiveness outcomes.* " In addition, we have chosen to include a positive example in the box, where the difference in baseline risk between the RCT and reality was taken into account. We have chosen to refer to a review and to state the main reason for the major differences (and not the underlying individual studies): "*In an HTA report comparing drug-eluting stents (DES) and bare-metal stents (BMS), the authors performed a review of the economic literature and identified rather opposite results: some authors indicate that DES may be cost-effective or even cost-saving in specific patients, while others mention DES is not cost-effective with ICERs of about 200,000 Canadian dollar per QALY gained.[103, 104] One of the most important determining variables for the ICER, next to the price difference of DES and BMS, was the baseline repeat revascularisation rate with BMS.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
76	ISPOR	39	see 3.4	13	It is challenging to establish updated baseline risk data for the relevant population at the national level. Register data will often be not adequately detailed. There is a risk that patients from the local/ national registeries will differ from pasients in the study.	minor	We added the following: "*We note that it can be a challenge to find up-to-date information on the baseline risk for a particular event at the national level. If such information is available, e.g. from administrative databases, the reliability must be checked.* "
77	PHARMERIT	40	3.4	Box 10	Text should say "where" not "were" - "The influence of this protocol-driven follow-up was also demonstrated in another study where the baseline risk for both MACE…"	Linguistic	This is corrected.
78	PHARMERIT	45	3.6	12	Does EUnetHTA have a position on whether the 3L or 5L EQ-5D is preferred?	Minor	At this point, as far as we know, EUnetHTA has not explicitly expressed a preference for the 5L or 3L. There are differences between the two (impact on ceiling effect, responsiveness/discrimination between health states, improvement in the description of the mobility domain, ...). However, this general guidance document is not the preferred document to discuss this specific issue.
79	EURORDIS	45	3.6	8_9	The following is presented as a "pitfall": "applying hypothetical (non evidence-based, e.g. evidence based on expert opinion or non-comparative observational studies) utility weights in the economic evaluations". That is exactly the case of cost-per-QALY method and the associates scale EQ-5D, which have been deeply and widely criticised. Though this method seems, nowadays, to have no valid alternative, we would not present it as the primary (neither absolute or objective) method. See, for instance: Collins, Latimer 2013, (BMJ 2013;346:f1363. https://www.bmj.com/bmj/section-pdf/187873?path=/bmj/346/7905/Analysis.full.pdf), focusing on end-of-life QALY gained and CE threshold, or Soares 2012, British Medical Bulletin 2012; 101: 17–31 (DOI:10.1093/bmb/lds003)	major	What we want to say is that no QoL was measured using a specific instruments and that the researcher assumes a particular utility that is not supported by any evidence. The second part of the following sentence clarifies this: "• *Applying hypothetical (non-evidence based, e.g. based on expert opinion or non-comparative observational studies) utility weights in the economic evaluations because no generic utility instrument is used in the underlying trials.* " In the case of the EQ-5D, researchers use one of the instruments for measuring the health state of the patient, after which they convert this health state to utilities via a tariff list.
80	EFPIA	45	3.6		Quality of life - The guidelines could focus further on how to address quality of life in babies / children and individuals who are not able to provide self reported responses. This is an important issue in diseaases such as Alzheimer's as the disease progresses. For example - what are the implications of moving from self-reported to proxy based outcomes? Furthermore, in diseases where progression can take up to 30 years (again AD), need guidelines in combining trial and not trial data.	Major	This falls outside the scope of this report with points for consideration when making a critical assessment of existing economic evaluations. The EUnetHTA guideline on HRQoL also contains a section on proxy measurements. We refer to this guideline at the end of this part: "• *Endpoints used for Relative Effectiveness Assessment: health-related quality of life and utility measures. Methodological Guideline: EUnetHTA; 2015.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
81	ISPOR	45	3.6		Quality of life - The guidelines could focus further on how to address quality of life in babies / children and individuals who are not able to provide self reported responses. This is an important issue in diseaases such as Alzheimer's as the disease progresses. For example - what are the implications of moving from self-reported to proxy based outcomes? Furthermore, in diseases where progression can take up to 30 years (again AD), need guidelines in combining trial and non-trial data.	Major	see ID80
82	ISPOR	45	3.6		The HRQoL refer to generic instruments and slighty on others with a focus on suggesting mapping exercises to generic PROs. We find in unfair to put forward the EQ5D as the almost only option for conducting economic evaluations. An economic evaluation is not always about cost-utility. It should also be a cost-effectiveness discussion, and this might be way more informative than the QALY discussion especially when EQ5D is not sensitive to certan disease attributes. The section should be expanded with proposals for improvement in that field. Lots of literature is published to indicate the limitations of generic PROs. What we should be communicating through this document is a way around EQ5Ds when there is strong belief that this not an appropriate and patient-centered tool for certain diseases so economic evalulation move beyond a QALY conversation and become more relevant to different stakeholders than NICE and ICER.	Major	We don't put forward the EQ-5D as the almost only option for conducting economic evaluations. We only note the observation from a previous systematic review: "*Based on the review of guidelines used by EUnetHTA partners, EQ-5D is the most commonly recommended instrument for derivation of HRQoL weights, although other instruments are also mentioned (e.g. HUI, SF-6D or 15D).* " We agree that the validity of the instrument is also of great importance. The mention the following in the first paragraph of this section: "*As for all applied questionnaires, the EUnetHTA guidelines also state that "documentation of the validity, reliability, responsiveness and acceptability of the HRQoL instruments used in REA should be provided."* It is not the intention of this report to list alternatives of what a researcher must / can do if it appears that the existing generic utility instruments do not prove to be valid.
83	INFARMED	45	3.6.		A major difficulty is about choosing between various generic measures. A recent debate emerged about the use of EQ-5D-5L versus EQ-5D-3L. Although I know there is no clear answer, I think that this document should say something about choice between generic measures.	major	see ID78
84	ISPOR	46	see 3.6	18	Different scenarios should be discussed: what if QALY results from the study show no difference between arms, but external QALY values based on literature search and used in the model show a difference which is clinically important?	major	There is no general answer to this. In some cases, it will be justified to model one situation as a base case and the other in a scenario analysis. In other cases, there may be arguments for clearly favouring one of the two situations over the other. Since no clear point for consideration can be given here that specifically applies to this, we prefer not to add this comment to our non-exhaustive list.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
85	EURORDIS	46	3.6	18_24	The following is presented as a “pitfall”: “applying hypothetical (non evidence-based, e.g. evidence based on expert opinion or non-comparative observational studies) utility weights in the economic evaluations”. That is exactly the case of cost-per-QALY method and the associates scale EQ-5D, which have been deeply and widely criticised. Though this method seems, nowadays, to have no valid alternative, we would not present it as the primary (neither absolute or objective) method. See, for instance: Collins, Latimer 2013, (BMJ 2013;346:f1363. https://www.bmj.com/bmj/section-pdf/187873?path=/bmj/346/7905/Analysis.full.pdf), focusing on end-of-life QALY gained and CE threshold, or Soares 2012, British Medical Bulletin 2012; 101: 17–31 (DOI:10.1093/bmb/lds003)	major	see ID79
86	ISPOR	46	3.6		Your discussion of "mapping" HRQL and utilities may be usefully complemented by discussions in a recent ISPOR Task Force report: Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to estimate health-state utility from non–preference-based outcome measures: an ISPOR Good Practices for Outcomes Research Task Force Report. Value Health 2017; 20(1):18-27.		We have added this reference in a footnote to the bullet about mapping: "*We refer to the ISPOR Good Practices for Outcomes Research Task Force Report on mapping for further information on this topic.* "
87	ISPOR	46	see 3.6		When is it advisable to refrain from disease specific QALYs for same condition?		This falls outside the scope of this report. Another EUnetHTA guideline deals with this. We refer to this guideline at the end of this part: "• *Endpoints used for Relative Effectiveness Assessment: health-related quality of life and utility measures. Methodological Guideline: EUnetHTA; 2015.* " In that EUnetHTA guideline, for example, the following is mentioned: "*REA performed for informing resource allocation decisions within indications can be based on validated comprehensive disease-specific HRQoL data, as comparability across indications is in this case less important. Nevertheless, the consideration of generic HRQoL data remains useful for reasons of coherence in the valuation of health benefits, and in consequence, transparency of the decision-making process.* "
88	ISPOR	46	see 3.6		Discuss pitfalls for QALY utilities obtained by regression/ multiple regression?		This falls outside the scope of this report about "Practical considerations when critically assessing economic evaluations"
89	ISPOR	46	see 3.6		Should the HTA control QALY- utilities against earlier publications?		This is a general comment that applies to many things. It is often interesting to see what the results are in other studies (whether it is QALYs or a different outcome). If there are differences, it is often interesting to find out where these differences come from.
90	PHARMERIT	47	3.6	29-34	As an adjunct to this paragraph, it may also be worth suggesting that utilities should be adjusted for age throughout the model time horizon (not just at baseline)	Minor	We added the following bullet: "• *Similarly, in the case of e.g. extrapolations to a longer time horizon, it is also important to check whether it is necessary to adjust the utilities for ageing.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
91	EFPIA	47		33-34	Using the utility values of general population for patient without AEs would probably need further clarification. I suppose the intention here is to say " age-adjusted utilies obtained from genral population for the hypothetical scenario with relevant condition/disease"? Othervise it needs to be additionally assumed that the condition which the intervetion is targeting does not have any disease specific symptions which would have impact on HRQoL.	Linguistic	We added the additional assumption between brackets in our text: "*In case the population of interest without adverse events is similar to the general population (i.e. without any disease-specific symptoms which would have an impact on HRQoL), age adjusted utility values of the general population could be used instead of applying a utility value of 1.* "
92	ISPOR	47		33-34	Using the utility values of general population for patient without AEs would probably need further clarification. I suppose the intention here is to say " age-adjusted utilities obtained from the general population for the hypothetical scenario with relevant condition/disease"? Otherwise it needs to be additionally assumed that the condition which the intervention is targeting does not have any disease specific symptions which would have an impact on HRQoL.	linguistic?	see ID91
93	ISPOR	50	see 3.7	28	Ideally we would like to see that numerical relationship between the size of incremental PFS and OS results is established. We need effect size to be expressed in numbers not only direction of the relationship.	major	we added the following footnote: "*We note that, ideally, researchers not only report the direction of the link but also its size* "
94	ISPOR	51	see 3.7	16-18	Could it be relevant for adjuvant therapy as well?	minor	The example of adjuvant therapy fits under usual circumstances in the list where there is a need for evidence to support the link between surrogate and final endpoints where possible. The other two examples in the text (very slowly progressive diseases or rare diseases, where long term clinical outcomes and/or big samples are not available) are of a different order. However, also in these cases, it is worthwhile to see which evidence is available to support or reject the hypothesis of a link between surrogate and final endpoints.
95	EURORDIS	51	3.7	21_30	The statement added in that note seems to question the Regulatory work as such, without context nor clear justification. The conclusion is moreover not justified as far as a primary endpoint as OS is not the only possible benefit we can measure in a cancer drug. Other possible benefit may weigh (and actually weigh) on the benefit/risk balance. In one case could be the administration, in another case the presence/absence of certain side effects, etc. We don't understand, in this context, the sens of a so general conclusion with so weak justification, on a so bigger issue, which may deserve to be adressed otherwise. We suggest deletion. About the general issue of Regulatory system see our general comments.	major	See ID28. This report applies to economic evaluations within the context of HTA and not to the benefit-risk assessment for the registration of medicines. It is important to be aware that the evidence that may be sufficient for registration is not always sufficient to formulate a positive conclusion and/or recommendation in an HTA report. That is why it is essential to include this note.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
96	EFPIA	55	3.8	20	The time horizon should be also specified in light of realistic elements concerning the potential population level treatment effect (i.e., herd immunity) going beyond long-term data measured in RCTs and observational studies and history of the disease. Thus, we suggest to add the potential population-level impact of treatment to the arguments of long-term data and history of the disease.	Major	This is indeed one of the many options with an incremental effect in the longer term. This falls under the general description that is included as follows in the text: "*the time horizon for the reference case analysis should be sufficiently long to reflect all relevant differences in costs or outcomes between the technologies being compared* ".
97	ISPOR	56	see 3.8	6	Does it mean that the time horizon should be made shorter? How much shorter? Are there any recommendations?	major	No, it is not intended to represent a consensus here about which time horizon researchers should use. The general observation remains that, in the first instance, researchers should follow the national guidelines. The sentence in the text is the following: "*economic claims based on models with very extended time horizons and predominantly extrapolated benefits will be less certain and are likely to be less convincing to the PBAC.* " This does not prevent researchers from choosing a specific time horizon as the base case and also displaying the results of scenario analyses with a different time horizon.
98	PHARMERIT	56	3.8.1	8	There is a lack of clarity on which "Parameters" are referred to.	Minor	This concerns all possible incremental parameters linked to incremental costs or benefits. We changed the text as follows: "• *Parameters related to incremental costs and benefits may vary across simulations ...* "
99	PHARMERIT	56	3.8.1	9-10	There is no elaboration on how simplifying temporal aspects of parameters supports the analysis of uncertainty. One may argue that this simplification leads to an underestimation of the uncertainty, which could otherwise be captured through e.g. Monte-Carlo simulation.	Major	This is indeed confusing. This is not about the parameter uncertainty that is modelled via probabilistic distributions and Monte-Carlo simulations (see section 3.12). This is about parameters that appear in the economic model in the long-term time horizon and for which researchers use today's info, while they are not sure whether that parameter will still have the same value in the future. We have changed the text as follows to better express this and also added a second example: "• *Parameters related to incremental costs and benefits may vary across simulations but be held at its current value over the time horizon within a simulation. This simplification supports the analysis of uncertainty regarding what we know now, rather than what is unknown about future events (e.g. how will population mortality rates look like in the future? Will the treatment cost for a specific adverse event be higher or lower in the future?).* "
100	EFPIA	63	3.9	1-21	The guideline does not refer to different discount rates between effects and costs. Especially in preventive interventions using different discount rates can be crucial, since effects becomming relevant in the far future (after decades) can be marginalized when using the same rates "paralizing paradox".	Major	Countries like Belgium and the Netherlands use a lower discount rate for benefits than for costs. It is not our intention to overrule national guidelines. We point this out as follows: "• *The discount rates stated in the national guidelines for economic evaluations should be applied.* • *Be aware of the possible large impact of different discount rates, especially in long-term models (see Box 20).* "
101	ISPOR	63	3.9		There are no comments on differential discounting of costs and benefits and whether or not this is appropriate. Thus, they recommend showing undiscounted results but I would suggest that they also recommend showing discounted results with the same discount applied to costs and benefits if this is not the recommendation of the local country to show the decision maker how much difference this makes to the results.		see ID100 Just as for all other elements in economic evaluations, we recommend that researchers always follow the national guidelines. Both the Dutch and Belgian guidelines recommend using an equal discount rate in scenario analyses. It is not the intention to draw up new guidelines in this report or to list all specific national guidelines. We refer to an overview including the different national discount rates in another EUnetHTA report: "*The EUnetHTA guideline on methods for health economic evaluations states that the impact of discounting in economic evaluation is often substantial.[1] Table A21 in this EUnetHTA guideline[1] provides an overview of discount rates for costs and effects in 24 countries. These guidelines should be followed in country-specific economic evaluations. In sensitivity analysis, different discount rates are applied and it is recommended that both the discounted and undiscounted results are shown.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
102	ISPOR	63	3.9		**First (p.10), has an error in the Section listing for the table on this consideration. For a topic that can lead to tremendous variability in modelling results, this section gives the lightest of coverage on the topic and although a political issue with the different country guidelines bodies, the topic of difference in discount rates used for cost and utilities in the same model can lead to tremendous divergence. This issue should be covered in depth here, as it would in academic courses on economic evaluations. Its absence is surprising in an economic guideline.	major	The cross reference on page 10 has been restored. The following is stated in the points for consideration: "• *Be aware of the possible large impact of different discount rates, especially in long-term models (see Box 20).* " The last sentence in this box also emphasizes this: "*Differences in the discount rate might thus have a considerable influence on the incremental impact on future costs and effects, and thus the ICER.* "
103	PHARMERIT	64	3.10	3	Check spacing between the section number and the section header - not consistent with other sections	Minor	OK, we will check this in the final version of the report.
104	PHARMERIT	64	3.10	-	The section highlights that different cost categories may need to be included depending on the perspective and also provides an example to this end. However, it does not discuss this for effects/utilities, e.g. the inclusion of caregiver utilities. Consider highlighting that the perspective not only has implications on chosen cost categories but also on the chosen effects categories and providing an example for this.	Major	The following sentence is mentioned in the report: "*The perspective is thus specified on both the effect and cost side, not only on the cost side.* " We thought this was an excellent place to add the example of the reviewer in a footnote. We have also added two recent references that further clarify this: "*This concerns, for example, whether or not to add the impact on the caregiver utilities. In a systematic review of literature on spillover effects on caregivers and family members, Alzheimer's disease and other types of dementia were the most frequent focus.{Wittenberg, 2019 #299} However, most Alzheimer's disease/dementia cost-utility analyses incorporated spillover costs, often as caregiver time costs, but considered spillover health impacts less often.{Lin, 2019 #300} If considered relevant, it is important to try to take this into account when setting up research protocols in order to gather reliable information on these spillover health impacts.* "
105	INFARMED	64	3.10.		The societal perspective may include very different resources, such as productivity costs, co-payments, informal care, transportations... It is important that these dimensions are clearly stated, defined, and presented separately. Also the State perspective may include costs for the sectors of education, social security, transportations, etc. The same comment applies.	major	Here too, the intention is not to draw up new guidelines ourselves, but to refer to the existing guidelines and to include points for consideration. We mention the following: "*National guidelines should be followed when performing an economic evaluation. Based on an overview of these guidelines, the EUnetHTA guideline on methods for health economic evaluations recommends the following:[1]* *• "Economic evaluations should at minimum be conducted from a health care perspective. However, several countries require a societal perspective. Presenting the use of resources as related to other sectors of society may increase the usefulness of the analysis to more EUnetHTA partners. Regardless of perspective taken, it is recommended that the use of resources is presented in as detailed a manner as possible. For example, if a societal perspective is used, indirect costs should be presented separately."* One of the bullets in the points for consideration also mentions the following: "• *The perspective defines the perimeter of consequences of health interventions to consider. It should be transparent which outcomes and costs are studied (e.g. only within or also outside the health care sector), whose outcomes (e.g. only for the patient or also for the caregiver or society) and which costs are studied (e.g. are only costs for the government included or also patient's co-payments and other costs).* " We believe that here too, it is important to look at the national guidelines to see if more details are given about what a particular perspective entails. We give the example of Belgium: "*For example, in the Belgian guidelines for economic evaluations, "the identification, measurement and valuation of costs should be consistent with the perspective of the Belgian health care payers."[92] In this guideline, 'health care payers' refers to both the patients, the federal government and the communities.* " We think it is redundant to add definitions for productivity costs, co-payments, informal care, and transportation.
106	PHARMERIT	65	3.10	2-5	Swap 1st and 2nd sentence of 1st bullet	Minor	We changed the order of these sentences.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
107	PHARMERIT	65	3.10	17-24	3rd bullet needs to be clearer and more informative for a researcher which may read this: which elements/parameters? Suggest to rewrite paragraph.	Minor	We have changed the first sentence to make this more clear. "• *For researchers, e.g. if a research protocol is drawn up for a clinical trial, it is important to think about the potential incremental impact on both costs and effects of using an intervention in comparison with a comparator. ...* "
108	PHARMERIT	65	3.10	17-24	Include an alternative for researchers not involved in protocol development: how should they go about this?	Minor	This is about researchers involved in drawing up a protocol. It doesn't make much sense to discuss this issue from the perspective of people not involved in setting up research protocols. This shows that it is essential to involve the right experts in time. We have therefore added the following sentence at the end of this section: "*Timely involvement of an expert with knowledge of economic evaluations can ensure that the right information is collected. Having such information at one's disposal might support the conduct and improve the quality of an economic evaluation when the trial is finished.* "
109	EFPIA	65	3.10	Box 21	Please offer an European example for applying more than one perspectives and not a Canadian one.	Major	We chose this example because it uses three different perspectives, which are well described, and the results are different depending on the perspective used. In the other examples we found, the conclusions were not different according to the perspective used, which we thought was less attractive as an example.
110	PHARMERIT	66	3.11	1	Check spacing between the section number and the section header - not consistent with other sections	Minor	see ID103
111	PHARMERIT	66	3.10	Box 21	This example would benefit from highlighting which specific costs were included for which perspective to further illustrate the section's point.	Minor	We added further information between brackets retrieved from the original paper: "... • the patient including costs incurred by the patient (transportation-related costs, cost of babysitter or housekeeper, and medical supplies not provided or reimbursed), and • society including all costs comprising productivity costs (time lost from work and activities for the patient and caregiver). ... "
112	PHARMERIT	66	3.10	Box 21	Inconsistent display of numbers" "$3090", "$12 107" and "$50,000". Consistently use one type of formating	Minor	This is corrected. We use the following notation throughout the report: 1, 10, 100, 1000, 10 000, 100 000.
113	PHARMERIT	66	3.11	NA	Resource use is only mentioned in one bullet. Internationally, HTABs require substantial substantiating of final costs by associated resource use. This topic is currently underrepresented. Consider adding in this chapter or dedicating its own chapter.	Major	It is outside the scope of this report to indicate how costs can be collected or presented.
114	ISPOR	67	see 3.11	3	Should unit costs be establish as marginal costs or average costs per unit/ resource? Any recommendations?	major	Here too, it is not the intention to write recommendations. As stated at the beginning of the document: "*In what follows we provide an overview of points for consideration for all the elements that are listed in Table 1. Where possible, we refer to existing recommendations from, for example, EUnetHTA guidelines. Then we give a non-exhaustive list of a number of points for consideration, followed by a number of examples that are presented in boxes.* " Concerning marginal or average costs: For the calculation of the ICER, we need to measure/calculate the incremental costs. Researchers can do this both by taking the difference between the average costs of two treatment arms or directly through focussing on the incremental costs. The marginal cost (in the meaning of "the change in the total cost that arises when the quantity produced is incremented by one unit") is not used in this context (where researchers usually make evaluations from the perspective of the health care payer or a societal perspective).
115	PHARMERIT	67	3.11	5	"countries" --> "country"	Linguistic	This is corrected.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
116	INFARMED	67	3.11.	13	Something should be said about unit costs. Sometimes, studies use costs from accountancy, which may not be appropriate to capture the value of the service, as it is based on accountancy rules. Sometimes, unit costs are based on prices, which may reflect incentives or other non-costs dimensions. In the societal perspective, calculation of productivity losses requires adequate estimates of the marginal product of labour, which is not straighforward.	minor	This is not within the scope of this report. We provide an overview of points for consideration. We do not repeat in detail what national guidelines do or do not tell about this. We state that it is a point of attention and provide an example. This happens, for example, here in "Box 23: Be aware of the financing system in different countries". The two following points for consideration also touch on this: "• *The cost items to be included should be determined by the chosen perspective. But administrative rules differ between nations, as healthcare is organized differently across countries. It is thus possible that costs that fall on healthcare authorities in one country are paid by some other entity, such as municipality or employers, in another country.* • *Be aware of possible differences in financing systems between countries when gathering cost information (see Box 23). Differences in financing might also impact the clinical pathway and related costs.* "
117	PHARMERIT	67	3.11	3-36	Consider including a bullet on costing catalogues like av	Minor	It is not the intention to give an overview of what all national guidelines state or which data sources these guidelines refer to.
118	PHARMERIT	67	3.11	8-17	Swap 2nd and 3rd bullet	Minor	We changed the order of these bullets.
119	PHARMERIT	67	3.11	8-12	Consider including a similar element like in the 3rde bullet of section 3.10: how should researchers take this into account as part of their trial development.	Minor	Information collected in trials should be published transparently afterwards. This also applies to reporting on cost information if gathering such information was part of the research protocol. We think it is redundant to mention this explicitly.
120	PHARMERIT	67	3.11	4-7,18-25	Consider merging bullet 1, 4 and 5 or grouping them together as the messages are very similar	Minor	we changed the order of these bullets so these items are mentioned next to each other.
121	PHARMERIT	67	3.11	8-12,30-35	Consider merging bullet 2, 7 and 8 or grouping them together as the messages are very similar	Minor	we changed the order of these bullets so these items are mentioned next to each other. This also changed the order of the three Boxes with examples.
122	ISPOR	67	3.11		lines 1 and 2 – if the costs are the same for both interventions they should not be of interest for budget impact analysis.	Minor	We deleted this sentence.
123	ISPOR	67	3.11		No guidance is given on how to deal with fixed versus variable costs – e.g. reducing hospital days might not change the hospital's fixed costs – at least not in the short run.	Minor	in section 3.11 under 'points for consideration' section we added the following point: "*Costs can comprise fixed (e.g., capital expenditure, salaries, building maintenance) and variable (e.g., medication, diagnostic and therapeutic supplies) components. In the context of health care, and particularly hospital care, a large part of costs accruing are typically fixed in nature. An intervention that achieves efficiency gains, such as through reduced patient length of stay, may have little or no impact on fixed costs over the short term. In reviewing an economic evaluation, one should be cognisant of how fixed and variables costs may be impacted by an intervention.* "
124	ISPOR	67	3.11		What about the opportunity costs of labor spent on a new intervention?	Minor	We refer to the national guidelines for these elements. We have added the following sentence at the beginning of this part: "*As in the previous section on the perspective, here too the national guidelines on costs must be followed. Based on an overview of these national guidelines, Tthe EUnetHTA guideline for methods of economic evaluations recommends that: ...* "
125	ISPOR	67	3.11		What about additional costs for other diseases when life expectancy is increased – e.g. for HIV infection – now CEAs could include costs of cardiovascular disease and other chronic illness not just associated with drug side effects but also with the aging of the population.	major	see ID124
126	PHARMERIT	68	3.11	Box 22	Consider adding a sentence explaining why this article was included as an example, i.e. highlighting different healthcare costs in EU countries.	Minor	We added the following sentence in the box: "*This study highlights the different healthcare costs in EU countries.* "
127	PHARMERIT	68	3.11	Box 22	In the 3rd paragraph, explain whether the cost items were pre-determined or not. This will be important for protocol developers.	Minor	While the document is intended for people performing economic evaluations rahter than for protocol developers, it is clear that the content of this box has implications for developing trial protocols. inclusion of per protocol costs may inflate the treatment cost relative to what is observerd in the real-world setting. However, it does not add much value in the context of this guidance document to state whether cost items were pre-determined or not.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
128	PHARMERIT	68	3.11	Box 22	Consider highlighting the importance of reporting resource use	Minor	We added the following sentence: "*This example also highlights the importance of transparent reporting of resource use to allow interpretation and potential adjustment of cost information.* "
129	PHARMERIT	69	3.11	Box 23	This box needs a clear explanation as to why this is important when considering costs in a health economic analysis. The last sentences should be further developed, including an elaboration on how this could be used in an CE analysis and how this can go wrong.	Minor	We added some further explanation at the end of this part: "*People not being aware of this financing system can make big mistakes when looking at the invoices to estimate costs of a specific hospitalization. As such, they would miss the part paid by the provisional twelfths and underestimate the 100% per diem price which might have a big impact on the incremental costs, ICER calculations, conclusions and recommendations.* "
130	PHARMERIT	69	3.11	Box 24	Consider discussing resource use data collection as part of trials, as this is more likely to be collected than just costs. Additionally, this would be more applicable to researchers plus it would set the correct standard moving forward.	Minor	Although this is very interesting, this falls outside the scope of this report.
131	EFPIA	70	3.1		Handling uncertainty: There is a need to add some guidance for the pragmatic side of evaluations as the basis of decision making. Specifically about dealing with the uncertainty that stems from limitations in data available. For example - 1) What assumptions should a decision maker accept with regards to long term effect when there is only trial length data available? 2) How much should the ICER threshold be lowered in relation to the degree of uncertainty if at all? Which/What measures should be used to categorize the level of uncertainty? On these questions any many other, those who are examining HE evaluations in order to advice decision makers are sometimes inconsistent and it is not always clear how they reached their preferred set of assumptions/recommendations. The methodological guideline would be even more useful if it looked critically into this side of the process as well.	Minor	The report does not aim to determine how a decision-making process works in different countries. It is impossible to map here in this EUnetHTA report "what assumptions a decision maker should accept", "how much the ICER threshold should be lowered", etc.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
132	ISPOR	70	3.1		Handling uncertainty: There is a need to add some guidance for the pragmatic side of evaluations as the basis of decision making. Specifically about dealing with the uncertainty that stems from limitations in data available. For example - 1) What assumptions should a decision maker accept with regards to long term effect when there is only trial length data available? 2) How much should the ICER threshold be lowered in relation to the degree of uncertainty if at all? Which/What measures should be used to categorize the level of uncertainty? On these questions any many other, those who are examining HE evaluations in order to advice decision makers are sometimes inconsistent and it is not always clear how they reached their preferred set of assumptions/recommendations. The methodological guideline would be even more useful if it looked critically into this side of the process as well.	Minor	see ID131
133	ISPOR	70	3.12		The extrapolation of long-term costs and outcomes is surrounded with uncertainty. It's important to properly address this structural uncertainty in scenario analysis. Perhaps this could be highlighted more in the guideline.		We added this sentence in our text: "*The extrapolation of long-term costs and outcomes is also surrounded with uncertainty. Exploring structural uncertainty can highlight where some assumptions may have a substantive impact on model outputs, and this is often usually carried out using scenario analyses.* "
134	ISPOR	71	3.12		I scanned this section very quickly but did not see anything about NOT using arbitrary ranges like + or – 20% without any rationale based on observed data – yet this is VERY common.	Minor	We have added this comment to the following sentence: "*Parameters defined with very narrow confidence intervals may be a cause for concern.* " The footnote added is the following: "*For example, it may be questioned why a uniform distribution is used in which the average is randomly changed by +/- 5% if there is evidence that the spread around the average cost is much wider.* "
135	ISPOR	71	3.12		I also did not see anything about ensuring that the variables included in the PSA were not correlated with each other – or if they were it is accounted for in the analysis	Minor	the following bullet is included in the text: "• *Have correlations between parameters been taken into account in the sensitivity analysis? While it is unusual for models to incorporate explicit correlations between parameter values, when they are included it is essential that those correlations are also appropriately reflected in any univariate sensitivity analysis or scenario analyses.* "
136	ISPOR	71	3.12		Also I did not see any guidance for structural uncertainty analysis about how to ensure that the CEA has properly tested all types of structural uncertainty – see this recent overview of the literature: Mauskopf J. Multivariable and Structural Uncertainty Analyses for Cost-Effectiveness Estimates: Back to the Future. Value Health. 2019 May;22(5):570-574. doi: 10.1016/j.jval.2018.11.013.	major	The following paragraph is mentioned in the text: "*To create a workable economic model, a variety of assumptions are generally made, such as how patients move between disease states.[173] These assumptions guide the structure of the model, and changing any of the assumptions will change the model structure and, potentially, generate different results. The extrapolation of long-term costs and outcomes is also surrounded with uncertainty. Exploring structural uncertainty can highlight where some assumptions may have a substantive impact on model outputs, and this is usually carried out using scenario analyses.* "
137	PHARMERIT	72	3.12	9	issue with the link to figure 15 which has meant an empty page and also missing text "Figure 15"	Linguistic	This is corrected.
138	ISPOR	72	see 3.12	Fig.15	The value of all parameters should be varied in a comparable manner -for ex. +/- CI or +/- 10%	major	We have presented the example as it was published. It is also not wrong to use a mix, e.g. apply the CI where a good sample is available and apply a wide range based on expert opinion if no good sample is available.
139	PHARMERIT	74	3.12	5	issue with the link to figure 14 which has meant an empty page	Linguistic	This is corrected.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
140	PHARMERIT	74	3.12	3-4	slight language change recommended here "A cost-effectiveness acceptability curve (CEAC) summarises the impact of uncertainty by showing the probability that an intervention is cost-effective for a range of willingness-to-pay thresholds". This is to avoid the confusion that the CEAC identifies the cost-effective option at any value of the threshold. It doesn't.	Major	This sentence is changed: "*A cost-effectiveness acceptability curve (CEAC) summarises the impact of uncertainty by showing the probability that an intervention is cost-effective for a range of willingness-to-pay thresholds.* "
141	EFPIA	75	3.12	1-5	Accepted approaches of expected value of perfect information and associated valued schould be layed down in the guideline.	Major	It is not our goal to inlcude an overview of what accepted approaches might be. Herefore, we refer researchers to the national guidelines (and not all national guidelines include a statement referring to EVPI). No major points for consideration related to EVPI were identified and thus not included in the report.
142	PHARMERIT	78	3.12	Box 25	issue with the link to figure 10 which has meant that it is included twice in the text	Linguistic	This is corrected.
143	PHARMERIT	82	3.12	Box 26	The statement " While it is very difficult to estimate the probability of an intervention being costeffective on the cost-effectiveness plane, this is much easier on the CEAC" is misleading. The CEAC presents the probability that the intervention is cost-effective for a range of values of the willingness-to-pay threshold - this is all it does.	Minor	The sentence has been changed as follows: "While it is very difficult to estimate on the cost-effectiveness plane the probability of an intervention being cost-effective for a range of WTP-values, this is much easier on the CEAC.
144	EURORDIS	87	3.13	16_18	*Check face validity of model results by comparing with the identfied evidence in the clinical part of HTA or with experts who know the disease and treatment under consideration,* **such as clinical and patients experts.** *Example: in haemophilia, doctors are confortable in assessing a bleed rate between 3 and 5 points as normal, while, for patients, one single point of difference represents a huge burden*	major	The sentence has been changed as follows: "• *Check face validity of model results by comparing with the identified evidence in the clinical part of the HTA or with clinical or patient experts who know about the disease and treatment under consideration.* "
145	EURORDIS	88	3.13	19	Not entirely sure what is meant by the word “traces” here. Does this have the same meaning as “accords” and if so would that be a better term to use?	Linguistic	traces is the terminology normally used to track patients through the model. We checked with the authors and did not identify a better word for this.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
146	ISPOR	91	3.14		It is not explained why transferability is even reasonable to do. Similar ICERs in different settings does not mean that results are transferable. This is far from reality. WTP levels, cost of teatment, cost of management, management of patients, system differences, cultural differences are not transferable, so how can an economic evaluation be? We would strongly recommend that the group does not push this elements forward as "next step" cause it can be really misleading especially in countries where lack of exertise on CEA/ CUA is obvious and decision-makers may blindly trust results from one setting to the other, across countries and within one country itself.	Major	With 'results' we refer to the results of the economic evaluation. We do not mean here the final policy decision that indeed depends on many other factors. We have clarified this in a footnote in the text to avoid misunderstandings: "*To avoid misunderstandings, we notice that we are referring here to the results of the economic evaluation (e.g. the calculated ICERs) and not to the results of the decision-making process of the policy makers, which can be influenced by many other factors.* "
147	ISPOR	91	3.14		Considerations and concerns around transferability are similar to those about economic evaluation in general. Results are country-specific and context-specific, and their use in decision-making must be viewed in that light. Uninformed application of economic results, from one's own country or another country, can lead to decisions that may be inappropriate for the decision context. Nevertheless, the underlying elements of an economic evaluation in one country can inform the evaluation process in another; it is at this level is where transferability considerations can be most useful.	major	This part is about the transferability of the results of the economic evaluation. We have clarified this in a footnote (see ID146)
148	PHARMERIT	92	3.14	20-27	Considering discussing cross-country collaborations such as EUnetHTA, but also BENELUXA	Minor	If there is cooperation, it is often limited to the clinical part of an HTA. Cross-country collaboration, which we do support, is not the topic of this report. That is why we do not think it is appropriate to add this here.
149	PHARMERIT	93	3.14	4	0 instead of appropriate cross-reference	Linguistic	We added the cross-reference.
150	PHARMERIT	93	3.14	5-7	In the context of transferability, it would be valuable to further explain the use of tariffs to derive country-specific utilities (based on e.g. EQ-5D questionaires). Consider including an example in a box, highlighting approaches which can be taken by countries for which no tariffs are available (and transferability is thus key)	Major	Adjusting to another value set is only possible if the individual patient data are available. We added the following in a footnote: "*Adjusting the QoL weights to a standard value set from another country is only possible if the health states are available at the patient level. This is often not possible as published clinical trial results on HRQoL are often only available as the average QoL (and CI) for the distinct treatment arms at different points in time.* " The choice of the value set for countries for which there is no value set available is described in the manuals of the respective generic utility instruments.
151	PHARMERIT	93	3.14	Box 29	Address hyperlink error	Minor	This is corrected.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
152	PHARMERIT	93	3.14	Box 29	The research outlined in this box indicates that there are very few systematic patterns in the economic evaluation of drugs. However, more recent research has shown that specific HTABs pay more or less attention to certain specific items (e.g. Tramonti et al. 2018. DIFFERENCES IN HTA COVERAGE RECOMMENDATIONS- COMPARING OPHTHALMOLOGY DRUG REIMBURSEMENT DECISIONS IN FOUR EUROPEAN COUNTRIES. ISPOR Spain 2018. Retrieved from: http://www.ispor.org/heor-resources/presentations-database/presentation/ispor-europe-2018/differences-in-hta-coverage-recommendations-comparing-ophthalmology-drug-reimbursement-decisions-in-four-european-countries; Nicod & Kanavox. 2012. Commonalities and differences in HTA outcomes: a comparative analysis of five countries and implications for coverage decisions. Health Policy. 108(2–3):), 167–177)	Major	The paper to which reference is made mainly concerns the reimbursement decisions that have been taken. This report deals with the results of the economic evaluations (and not the final policy decision).
153	ISPOR	94	3.15		There is no reference that ICER thresholds are arbitrarily set. This is a crucial point that lots of stakeholders struggle with, including economic assessors, policy makers, HTA bodies. Not sure why this document does not try to address the caveats of the pre-defined thresholds, the purpose of these, the lack of efficiency in using them to solve the actual problem to help the scientific community move forward and evolve. It would be great for this section to touch upon this before outlining the points of consideration which are all valid	Major	This would require an entirely separate report and falls outside the scope of this report. We limit ourselves to the following: "*There are different ways to set a cost-effectiveness threshold.[202, 206, 207] In this chapter we do not present, recommend or criticize any of them. We want to draw your attention to some issues when reading the interpretation of the results in an economic evaluation.* " We also explicitly mention that "• *There is often no explanation/justification for the selection of the cost-effectiveness threshold (or range of thresholds) (see Box 30).* "
154	INFARMED	96	3.15.	15	It is obvious that other criteria are used to take decisions, beyond cost-effectiveness. Yet, it is much less obvious that economic evaluations should report these other criteria, which health economists would hardly evaluate. I would clarify this point.	minor	In this point for consideration we only want to notice that there is more than just the cost-effectiveness argument. It is not our intention to say that a researcher should report these other criteria. Our focus in this report is on the economic evaluation itself.
155	PHARMERIT	96	3.15	27-28	Consider rewording for clarity	Linguistic	We changed this sentence as follows: "*It might be more difficult to interpret or discuss the results in function of different cost-effectiveness thresholds if no CEAC is presented (see Box 31).* "
156	INFARMED	97	3.15.	box 30	The box suggests that ICER thresholds are purely arbitrary, while efforts have been developed to measure it, e.g., on the basis of opportunity costs (Claxton et al, HTA 2015, Woods et al, Value in Health, 2016), or on the basis of willingness to pay (Shiroiwa et al, Health Economics, 2010).	major	There is much debate about this. One of the other reviewers, for example, states that "There is no reference that ICER thresholds are arbitrarily set." It is not the intention here to give an overview on research about the cost-effectiveness threshold. We only mention several points for consideration related to this topic (and one of them is that there is often no explanation/justification for the selection of the cost-effectiveness threshold).
157	PHARMERIT	100	3.15	Box 31	"Error! Reference source not found."	Linguistic	This is corrected.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
158	PHARMERIT	100	3.15	Box 31	There is no accepted approach to labelling of the quadrants of the CE plane, this the use of the term quadrants 1 and 2 is problematic. In addition, the explanation would be clearer if it stated that these quadrants involve the intevention being more effective than the comparator rather than "better" which is open to interpretation.	Minor	We changed the text as follows: "*Having 58.3% of the simulations in the first (north-east) and second (south-east) quadrant of the cost-effectiveness plane (i.e. the intervention is more effective than the comparator) also means that 41.7% of the simulations indicate worse results.* "
159	ISPOR	101	box 32		The tone of this section could be polished a bit. Use of phrases like "the authors even refer" or "Chappell et al even applies" makes it sound like the writers are trying pretty hard to make a point.	minor	We deleted a couple of times the word 'even'.
160	PHARMERIT	102	3.15	Box 32	Suggest textual change to " Using ICER threshold values that are too high equates to ignoring the economic argument in decision making".	Linguistic	we changed the text as follows: "*Using ICER threshold values that are too high rather equates to ignoring the economic argument in decision making.* "
161	PHARMERIT	102	3.15	Box 32	The last statement "When there is no explicit ICER threshold value and decision makers consider the presented ICER threshold too high to apply systematically, presenting the results on the CEAC provides a good alternative. This allows decision makers to interpret the results applying their own willingness/ability to-pay instead of the (possibly unrealistic) values that might be stated by authors of an economic evaluation." should include the clarification that, for a risk neutral decision-maker, the CEAC does not identify the cost-effective option at any value of the threshold but rather presents the uncertainty of the technology being cost-effective for any value of the threshold.		See ID140 & ID143.
162	PHARMERIT	107	5.1.1	4-5	Fix spacing	Minor	This is corrected.
163	PHARMERIT	107	5.1.1	7-8	Was a search protocol for "guidelines that were identified on websites" developed? Would be worthwhile to list the included sources (e.g. were ALL website of HTAB bodies globally considered or only a selected few?) and search terms to demonstrate impartial and transparent approach.	Minor	We added further explanation on our approach in the text: "*The main source of information were thus the existing guidelines that were identified on websites (EUnetHTA, websites of HTA bodies and ISPOR) and the experience of the authors, complemented by the comments of the external reviewers. The emphasis was on HTA bodies that are part of the EUnetHTA collaboration. As this document is intended for guidance and is not prescriptive, the review of existing guidelines was not systematic, but rather intended to identify areas of best practice and of common issues in the conduct of economic evaluations.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
164	PHARMERIT	125	5.2	15-16	The interpretation of the CEAC is incorrect. The CEAC gives the probability that the intervention is cost-effective BUT where the decision-maker is risk-neutral and the aim is to maximise health subject to a budget constraint the cost-effective option at any threshold should be determined by comparing the expected value of the ICER to the threshold (or for INB - either monetary or health - by comparing to 0) not by determingin which option has the highest probability of being cost-effective. As such the threshold where the cost-effective option switches between option C and option A is given by the ICER and not by the point where the CEACs cross.	Major	We changed the text as follows: "*Interpretation: The probabilistic results show that intervention C has a probability of more than 50% of being cost-effective up to a €300 000/QALY threshold. The probability of being a cost-effective intervention is more than 80% for a willingness to pay under €250 000/QALY.* "
165	PHARMERIT	126	5.3	NA	In this section it would be useful to include a small paragraph highlighting the open source R package *heemod.* This package is used for writing health economic Markov models in R, and is an example of how a global model engine (validated by open source community) can be used to generate a wide variety of different models. This is important when it comes to sharing, as only limited set of information needs to be shared in order for someone to reproduce the model using *heemod* themselves. More information here: https://arxiv.org/pdf/1702.03252.pdf. Figure 8 in particular demonstrates the user-adjustable (sharable) elements vs core engine functionality. The following link is to a paper which used heemod for its modelling, and all information needed to recreate the model is given in the supplementary appendix: https://jamanetwork.com/journals/jamaoncology/article-abstract/2716813. There are probably more like this.	Minor	The concept of model sharing is a general concept, which is also possible with other software. The easiest way is to make the model available. If this happens with software that everyone knows (such as Excel), then this has the advantage that everyone can view the model (while not everyone is familiar with R). We, therefore, prefer to talk here in general about model sharing, where we provide five examples in boxes without explicitly referring to specific software.
166	EURORDIS	0 general	3.2		One of the fundamental points of this paragraph is that the best clinical practice can also not correspond to the standard of care (reimbursed) or be far from the efficiency frontier (based on cost-effectiveness ratio).	minor	this is indeed described in the following sentence: "*Researchers should thus look further and also consider current standard practice or the reimbursed alternatives, which in some cases might be different from the optimal care described in practice guidelines.* " So we did not change this text.

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
167	EURORDIS	0 general	3.6		This chapter in particular, as other ones of this paper, seems to accepts the QALY and the cost-per-QALY outcome measure as the primary one, regardless of the concern that could be raised about this method. The fact that the same assumption is made in the EUnethTA guideline doens't minimize the issues. (see p. 45 L. 8-9 "the primary outcome measure(s) should be presented where appropriate as natural units (including life-years) and as QALYs".)	major	this is correct. We indeed refer to the findings of the EUnetHTA guideline on HRQoL. However, the paragraph also mentions the following nuance: "*As for all applied questionnaires, the EUnetHTA guidelines also state that "documentation of the validity, reliability, responsiveness and acceptability of the HRQoL instruments used in REA should be provided."[10]* "
168	EURORDIS	0 general	3.15		In addition to the main references of this chapter, such as WHO 2016 (*Cost-effectiveness thresholds: pros and cons*); Neyt et al. 2018; and KCE 2008 (*Threshold values for cost effectiveness in health care*), we suggest the OHE report of March 2018 (Towse, Garau, *Appraising Ultra-Orphan Drugs: Is Cost-Per-QALY Appropriate? A Review of the Evidence*), which includes conclusions such as: "Given the importance of non-QALY elements in the assessment of HST, such as treatment impact on the process of care and on the patients' or their carers' ability to go to school or to work, and issues in measuring quality of life when the population affected are infants or young children, it is inappropriate to focus the HST appraisal only on a cost-per-QALY measure". "Given the lack of empirical basis, the new £100,000 cost-per-QALY threshold and its further possible uplift by a factor of three seem arbitrary".	major	The reference the reviewer suggests is published as a 'review of the evidence'. However, no transparent search strategy is provided. The situation also focuses on the UK Highly Specialized Technology (HST) program, which is specific to the UK. The reference to which the reviewer refers also contains a number of statements that are not sufficiently nuanced. For example, in the introduction: "*NICE does not consider other costs, for example, those incurred by patients, family and carers, or by employers from having ill and absent workers.* " While the NICE guidelines state that these elements can ensure that a different threshold is applied. The following is stated in the 2013 (and later) NICE guideline: "Above a most plausible ICER of £20,000 per QALY gained, judgements about the acceptability of the technology as an effective use of NHS resources will specifically take account of the following factors: ... Aspects that relate to non-health objectives of the NHS". The report the reviewer refers to also mentions several challenges with QALYs. However, this report is not an evaluation of the pros and cons of working with QALYs. Furthermore, no country applies the ICER calculation in €/QALY as the sole decision criterion. It is one of the elements that decision makers can take into account. In our report, we mention the following: "• *In all countries decision making is not solely based on cost-effectiveness considerations. The technology is assessed based on efficiency criteria together with other criteria. In the presence of high ICERs, those other criteria become more important."* This reflects the remark that it is inappropriate to focus an appraisal only on a cost/QALY measure. The first two points for consideration are the following: "• *Cost-effectiveness is not the only criterion to make decisions. It is not just because an intervention has an acceptable ICER that it will be reimbursed and vice versa. Other elements like the uncertainty around the estimates, the budget impact and budgetary context, the degree of unmet medical need, etc. also influence the reimbursement decision.* • *There is often no explanation/justification for the selection of the cost-effectiveness threshold (or range of thresholds) (see Box 30). An explanation/justification for the selection of the applied threshold should be given. Readers should pay caution if authors refer to the ICER of an intervention that received a positive reimbursement decision since it is possible that the economic criterion has been ignored/overruled in this decision (e.g. because it was a very small population with a very severe disease and high unmet medical need).* " We think this adds sufficient nuance in our report.
169	EURORDIS	0 general	3.16		Conflict of interest is another concept that is widely misunderstood and under-recognised. As a result of this we see new bias introduced into analyses resulting from efforts to eliminate bias.	major	We added the following in the points for consideration: "• *We must be careful that efforts to eliminate bias do not lead to the introduction of new bias. The critical assessment of identified studies is preferable to the complete disregard of studies by authors with a possible CoI."*

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
170	EURORDIS	0 general	3.16		The first thing to recognise is that bias exists everywhere and it is not just limited to the obvious causes, such as monetary incentives provided by industry etc in the form of direct payments, research grants and etc. Professional reputations are often more important than financial gain as a source of bias but is much harder to police – seemingly unconnected "independent" advisors to an assessment may have conflicts they are not even aware of and most assessors will have their own inherent bias for or against the approach under consideration.	major	The methodology of evidence-based medicine that is applied in HTA reports supports an assessment that is as objective as possible. We cannot support the reviewer's statement that "most assessors will have their own inherent bias for or against the approach under consideration"
171	EURORDIS	0 general	3.16		It is also important to recognise that even where the **potential** for bias is identified and recognised, that in itself is not the same as proving that the bias has influenced the evidence. Thankfully many (not all) experts are aware of their own potential for bias and make strenuous efforts to overcome that – sometimes even overcompensating.	major	see ID169
172	EURORDIS	0 general	3.16		There is no mention about the potential conflict of interest of the assessors/evaluators, neither from authorities nor at decision-making level. In HTA, the public authorities, especially when the are also payers, represent one of the (financial !) interest in this field. That's an interest that could generate conflict. So, any very financial interest (private or public) should be assessed. Therefore the danger of bias and conflicts of HTA and public authorities experts should be examinated as carefully as the one of other external experts (like patients, for instance, which are often seen as an obstacle to quality and unbaised transparency).	major	This report deals with the economic aspect of an HTA, in support of, for example, reimbursement decisions. The scope does not include aspects related to the decision process and/or the decision of the policy maker that is influenced by many other factors. The potential CoI of the decision-maker is not part of economic evaluations and therefore falls outside the scope of this report.
173	EURORDIS	0 general	3.16		Additonal consideration: where there is no public access to both the information and the decisions, there is no transparency. Keeping distance between patient representatives and authorities is equal to avoid public scrutiny on the proces and the decision	major	This report is not about the transparency of the final policy decision. We focus on different elements of an economic evaluation. Nevertheless, transparency is indeed necessary to allow critical evaluation. In the introduction, we have added the following: "*Reporting guidelines are very helpful for both researchers writing down the study results of their economic evaluation, as well as assessors identifying the relevant elements when reading such studies. In fact, transparent reporting of the input variables and the assumptions made is necessary to enable a critical evaluation.* " This is also covered in the section on transferability: "• *It is impossible to assess transferability and it is difficult to transfer results to another setting if economic evaluation data and methods lack transparency.* "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
174	EURORDIS	0 general			Nonetheless, there are two missing factors, both in this chapter and in the aforementioned studies (including OHE 2018), related to severely disabling and/or some rare conditions. Firstly, the same principles stated in Section 3.4 'Baseline Risk' and 'Section 3.9 Discount rate' Box 20, should apply in reverse to the relative benefit of a technology and have an impact on the cost-utility-analysis as well as on the relative ICER Threshold. In fact, a gain which appears 'moderate' for an average patient/population, could be estimated as a major one for patients that start from a condition of severe disability. In that case, a standard QALY scale would fail to appreciate the benefit, unless it is adapted. Secondly, as a consequence, another missing factor is the willingness of society to pay for those who start from a condition of greatest disadvantage, due to their baseline condition, for which a 'moderate' improvement would weight as a major one.	major	There are no standard accepted adjustment scales for this (not for the QALYs nor for the willingness of society to pay). This does not mean that the policymaker cannot take this into account in his/her decision. The extent to which the policymaker takes this into account or not falls outside the scope of this report.
175	PHARMERIT	12-14	1	1-33,1-33,1-4	Consider clarifying whether this methodological guideline applies to pharmaceutical products, diagnostics, medical technology, given the example use in box 33 of section 3.16	Major	We added the following in the introduction: "*This not only applies to economic evaluations of pharmaceuticals, but also of other interventions such as medical devices, diagnostics, prevention or screening campaigns, etc.* "
176	INFARMED	23-24	Box 3		The HRT example is very enlightening, but I suggest to briefly detail the source of bias in observational studies, i.e., which characteristics differed between RCT and NRS, which may explain the discrepant results. The same for the Digoxin example.	minor	We added the following for the digoxin case: "In this case, digoxin is particularly prone to prescription bias as clinicians have been trained to use digoxin in sicker patients.[48] When these patients died, there was therefore a true but misleading association between death and digoxin." For the HRT case, we refer in a footnote to a study performed by Hartz et al. trying to identify the confounders that could explain the differences between the observational and RCT results: "Researchers tried to identify the conditions for valid observational studies (OS).{Hartz, 2013 #301} They studied the WHI data containing "*information on more than 800 possible confounders including information that made it possible to accurately predict HT [hormone therapy] use. It also contained information on factors that might have influenced response to HT. Some of these factors were related to the timing hypothesis (e.g., age, time since menopause, previous HT use, beginning HT after baseline), and some were identified empirically (e.g., blood pressure, previous coronary revascularisation and private medical insurance). Since OS and RCT participants differed with respect to these factors, these factors could have conceivably contributed to differences between the OSs and the RCTs. However, after taking into account all of these confounding factors and stratifying on factors that may have influenced the response to HT, OS and RCT differences remained."*{Hartz, 2013 #301} In their conclusion, the researchers state that they "*did not find that the comprehensive data provided by the WHI were adequate to overcome problems often attributed to OSs. The findings do not imply that most OSs are invalid. They do suggest, however, that given the current methodology, even very good OS datasets may not be adequate to give reliably valid results. … Without better OS methodology there will be underuse or misuse of OSs for comparative effectiveness research."*{Hartz, 2013 #301} "

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
177	EURORDIS	33_34	3.2		We would clarify further the lesson brought by the "non-medical example". The right comparator is the one considering different criteria and needs. That is why a four place pushcart, though it's a cheaper alternative and -finally- an effective one, would not fall into the potential comparators. Those criteria and needs can be best understood and selected via a discussion gathering different expertise and viewpoints.	minor	We only refer to the example as it was provided in the original paper. The comment that the relevant comparator can be collected through different expertise and viewpoints is displayed by adding the following in the text: "*Researchers should thus look further and also consider current standard practice or the reimbursed alternatives, which in some cases might be different from the optimal care described in practice guidelines. Expert opinion or patients' view might also be helpful in identifying relevant comparators.* "
178	EURORDIS	44_45	3.5	Box 11	"*In the absence of local data, there is a preference for relevant international data – which has been used by a second study* " This is a sensitive point for small countries, as they are unlikely to host clinical trials. In those countries, it is also hard, for assessors, to even get data from NHS system about their local patients. For example, in bleeding disorders, how many times a patients end up in hospitala and how much does it cost to the healthcare system. That is why small countries should be be encouraged to gather and publish their own local data.	minor	How data is gathered on a national level and how to access such data is out of the scope of this report.
179	EFPIA	64-66	3.10	13-14	On page 64 lines 13 and 14 it is stated that "The perspective is thus specified on both the effect and cost side, not only on the cost side." We agree in particular with the half-sentence of this statement but feel the "outcomes side" could be discussed more explicitly in this section as most of the arguments are on the costing side also including the example in the box 21. Thus, we recommend to either emphazise or potentially even include an example where for example quality of life loss in carers has been considered.	Major	The following footnote with two recent references were added in the text: "*This concerns, for example, whether or not to add the impact on the caregiver utilities. In a systematic review of literature on spillover effects on caregivers and family members, Alzheimer's disease and other types of dementia were the most frequent focus.{Wittenberg, 2019 #299} However, most Alzheimer's disease/dementia cost-utility analyses incorporated spillover costs, often as caregiver time costs, but considered spillover health impacts less often.{Lin, 2019 #300} If considered relevant, it is important to try to take this into account when setting up research protocols in order to gather reliable information on these spillover health impacts.* "
180	PHARMERIT	67,70	3.11	4, box 24	Consistently use either "healthcare" or "health care"	Linguistic	we changed this (except in references or quotes were healthcare was written) and systematically used health care.
181	PHARMERIT	70-85	3.12		There are a number of issues with the links and text in this section which mean vast amounts of text are repeated this makes it impossible to follow in it's current form.	Major	There was indeed a problem with crossreferences repeating a big part of the text and figures. This has been solved.
182	EFPIA	74-85	3.12		Some figures are unnessecarily doubled.	Minor	see ID181

ID	Organisation	Page number* (or 0)	Section number	Line number* (or range)	Comment and suggestion for rewording	Character of comment	Author's reply
183	PHARMERIT	92-94	3.14	-	Consider including an example on the use of other country's/EUnetHTA's health technology assessment in another country. For example, some Eastern European countries specifically take NICE assessments/conclusions into account as part of their assessment procedure. This, in order to help the reader understand the practical application of transferability.	Minor	The focus of this guidance document is about critically assessing economic evaluations. We prefer to refer to the HTA adaptation toolkit for this topic. The following is mentioned in the report: "*Within a previous EUnetHTA project, an HTA adaptation toolkit has been set up as an aid to HTA agencies in the adaptation of HTA reports from one setting into another. This toolkit also contains a list of relevance, reliability and transferability questions to ask when considering the adaptation of information and/or data on economic evaluations. We refer to this toolkit for an overview of these questions (see box 10 in Section 5.4 of the HTA adaptation toolkit, available on https://www.eunethta.eu/eunethta-hta-adaptation-toolkit/).* "
184	EURORDIS		3.4		See section 3.15 comments		