EUnetHTA 21 Public Consultation
Of D4.3.2 – Methods Guideline on comparators and comparisons

Name organisation & abbreviation	Country
European Union of General Practitioners/Family Physicians UEMO	Belgium
BIOTRONIK SE & Co. KG	Germany
Ecker + Ecker GmbH (E+E)	Germany
SKC Beratungsgesellschaft mbH (SKC)	Germany
Verband Forschender Arzneimittelhersteller (vfa) e.V	Germany
GKV-Spitzenverband, GKV-SV	Germany
German Medicines Manufacturer´s Association (BAH)	Germany
Lymphoma Coalition - Lymphoma Coalition Europe (LCE)	France
Bundesarbeitsgemeinschaft Selbsthilfe von Menschen mit Behinderung und chronischer Erkrankung und ihren Angehörigen e.V. (BAG SELBSTHILFE)	Germany
European Confederation of Pharmaceutical Entrepreneurs (EUCOPE)	Belgium
Edwards Lifesciences	Europe
European Federation of Pharmaceutical Industries and Associations (EFPIA)	Belgium
GSK	UK
IGES Institut GmbH and HealthEcon AG	Germany
F. Hoffmann-La Roche Ltd (Roche)	Switzerland
Lumanity	Lumanity is a global company with several European entities, including in Ireland and the Netherlands.
Advanced Medical Services	Germany

Please add extra rows as needed.

GmbH - AMS	
Bayer AG & Bayer Vital GmbH	Germany
Medtronic	Switzerland
AstraZeneca (AZ)	Global (UK based)
European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) HTA SIG	Europe
Les Entreprises du Médicament, Leem	France
Norwegian Institute of Public Health (NIPH)	Norway
Alliance for Regenerative Medicine (ARM)	Belgium
Takeda Pharmaceuticals International AG	Brussels, Switzerland, local operating companies across the European Union
MedTech Europe (MTE)	Europe - Belgium

Please add extra rows as needed.

General Response:

We thank the reviewers for their submissions. Given that many similar comments were made, we will try to address at least some of the main themes in the clarifying text below. Where more specific comments were received, we address them in the table of comments.

First, we wish to clarify that this guideline is primarily intended for assessors who will be reviewing the evidence provided by HTDs, even if we recognize that the methodological guidance will also be relevant for other users. Further there is an additional complimentary document called a Practical Guideline (D4.3.1) which will also support HTA assessors in writing up a report which has included a direct and/or indirect comparison. This guideline is neither a comprehensive statistical textbook, nor a statistical cookbook prescribing how to perform analyses depending on a given data situation. Rather, it summarizes the basic principles, strengths, and limitations of the many methods that can be used when combining evidence to complement the practical guideline. We have appropriately described the limitations when certain methods are used under certain conditions. We have deliberately not included value judgements in the guideline as per the HTA Regulation. The regulation does stipulate that assessors should describe the limitations around the evidence and, with this in mind, we have produced a guideline to enable this to be done.

This methodological guideline does highlight that well conducted randomized clinical trials provide the most robust data. This does not imply that only RCT data is acceptable.

With regard to evidence collected in the real world setting, the guideline describes situations where data are not randomised. The term real world evidence (RWE) or real world data (RWD) is loosely used to imply outside of clinical trial conditions or sometimes as outside the randomised conditions. However, this terminology is not helpful and evidence gathered in this manner should be identified by its design (see Pacheco, R.L., Martimbianco, A.L.C. & Riera, R. (2022): Let's end "real-world evidence" terminology usage: A study should be identified by its design. J. Clin. Epidemiol. 142, 249-251). The terms RWD or RWE do not describe the defining characteristics of the design of a study. Therefore, while we have not specifically included a section under these terms, they can be considered as included under the section for non-randomised evidence.

Many comments consider that in places the guideline is too strong in its wording or too weak in its recommendations. We have tried to strike a balance in highlighting the limitations that arise when certain conditions apply and the approaches that can be used when certain evidence is available. The degree of uncertainty which is acceptable may differ between Member States and therefore it is not the goal of this guideline to impose acceptability thresholds but to highlight the limitations that would arise with methods were certain assumptions not met.

The EUnetHTA 21 D4.3 Hands-On Group

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Dr Daniel Widmer UEMO	General	-	Very interesting and rigorous technical document. GPs wonder about comparisons concerning the practical use of technologies that would sometimes require a qualitative approach or mixed methods.	Thank you for your comment. Methods on qualitative and mixed methods were outside of the scope of the current methodological guidance.
BIOTRONIK SE & Co. KG	general		The document sets out to describe evidence synthesis via direct and indirect comparison methods. Unfortunately, the document neglects real-world evidence at the outset and therefore denies by default the inclusion of long-term data for any type of medical technology, be it a pharmaceutical, medical device, or IVD. Adding a section on the utilization of real-world evidence would add greatly to the guidance and future-proof it. As it currently stands, it will rarely be relevant for analysing evidence typically available for class IIb or class III medical devices and does not provide sufficient methodological guidance for any type of comparison including long-term outcomes data from e.g. registries.	We don't agree that this document neglects real world evidence. Please see the general comment at the beginning of this document which should provide clarity to the issues raised in relation to RWE.
Dr. Thomas Ecker, Ecker + Ecker GmbH	general		Methodological guideline D4.3.2 is an excellent summary of current HTA methodology on direct and indirect comparisons.	Thank you
Dr. Thomas Ecker, Ecker + Ecker GmbH	general		According to HTA Regulation (EU) 2021/2282 "*Member states shall (…) not request at the national level information, data, analyses or other evidence that has been submitted by the health technology developer at Union level in accordance with Article 10(1) or (5)"*. However, this proposition becomes potentially irrelevant once MS have methodological requirements on a MS-level which are different from the methodological requirements for EU HTA. Hence, guidance is needed how to resolve (potentially) conflicting requests between EU HTA and national (MS) HTA bodies. To give an example: Evidence synthesis conducted via a Bayesian method using a specific prior is accepted in EU HTA. However, national HTA bodies of member states consider the prior insufficient and require evidence synthesis conducted using a different prior or frequentist methods.	Thank you; however, as national methodological requirements are outside the scope of this deliverable, we do not propose to make changes in response to this comment. All aspects of implementation are likely to be addressed under the future Coordination Group (CG) governance structure
Prof. Matthias	general		Comment:	Thank you, we will adjust this

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
P. Schönermark, M.D., Ph.D. and Svenja Sake, Ph.D. (SKC)			According to the EUnetHTA project timelines as well as the content and cover page of the draft in question, sub-deliverable D4.3.2 is concerned with "direct and indirect comparisons" not "Guideline on comparators and comparisons" as written in the left column of this comment form. Concurrently and again according to the EUnetHTA project timelines, sub-deliverable D4.3.1 will be called and concerned with "Comparators and comparisons" which is wrongly referenced as "direct and indirect comparisons" throughout the present draft of sub-deliverable D4.3.2. We recommend to review and correct the references to the sub-deliverables D4.3.1 and D4.3.2 both in this comment form as well as in the D4.3.2 draft.	typographical issue.
Prof. Matthias P. Schönermark, M.D., Ph.D. and Svenja Sake, Ph.D. (SKC)	general		Comment: Overall, the statements in the document are rather unspecific: There are only few prioritizations of analysis options, concrete decision-making guidelines and specific consequences of certain evidence levels and analysis qualities. This likely results in uncertainties both for the HTD as well as the assessor and may limit the quality of the submitted information as well as the reliability and comparability of assessment procedures. We recommend to include a summary figure e.g., in the form of a decision tree which shows under which conditions (data availability, quality and robustness) which analyses are to be preferred or disregarded.	Thank you for this suggestion however it is not the purpose of this guideline to provide a decision framework as you describe. Please see general comments in response to comments received.
Sebastian Werner vfa	general		The proposed scoping approach in D4.2 of consolidating a PICOS survey from 27 member states will likely result in multiple comparators and potentially subpopulations requested for the JCA. With this approach, conducting indirect treatment comparisons will become a critical part of a JCA. The **requirements** for an acceptable ITC as set out in this draft guidance **are stringent and unfeasible for many new technologies**, particularly those included in the first JCA waves (orphan diseases, oncology and ATMPs). The combined impact of the two proposed guidance documents risks the JCA not being useful for MS for decision making.	Thank you but this is outside the scope of this guideline. The procedural aspects of the process will be considered under the governance of the future CG.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			The guidance on ITC and NMA should **allow for flexibility** to select the appropriate method for the decision problem in question. To ensure the best possible preparation, a **Scoping Meeting** should be held with participation of the HTDs, JCA assessors and clinical and patient experts where the PICOS, available evidence and therefore available methods of comparison should be discussed.	
Sebastian Werner vfa	general		**All sources of available evidence should be acknowledged**. The corresponding evidence level of the sources should be considered. The guidance should allow for flexibility to select the appropriate method for the decision problem in question. Instead of complete rejection of existing evidence due to violation of strict criteria the general approach should be to **use the best available evidence**. Complete exclusion of existing evidence introduces another risk of selection bias and should therefore be avoided in the context of an evidence-based approach. Methodologies for performing JCA should be adapted to include **specificities** of new health technologies for which some data may not be readily available. This may be the case for, inter alia, orphan drugs, vaccines, and ATMPs (HTAR [Article 4(1)]). Such adaptations are missing in the guideline and hence need to be addressed. There are methodological advancements to improve evidence generation in particular for small populations like the EU funded INSPIRE, that should be adequately acknowledged. Furthermore, due to practical and ethical considerations (e.g., challenges of patient recruitment for orphan diseases or in small and vulnerable populations) many medicines are developed and approved based on single arm trials; this is more likely to occur in therapies planned for Phase I & II of JCA (oncology, ATMPs, orphan drugs).	Thank you for your comment. Please refer to the general comments provided which explains the purpose of this document.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			To ensure EU patients have access to these novel therapies, it is imperative that appropriate novel approaches (such as MAICs, external controls etc.), which are based on the best available evidence, will be accepted as a means of assessing the comparative effects vs the full comparator basket.	
Sebastian Werner vfa	general		The document contains **several strong statements** indicating the "strong" limitations of various methods – suggesting that the default is to reject these methods. Instead, the default should be to apply innovative methods in order to use the best available evidence and to acknowledge and describe potential limitations, e.g., by conducting various sensitivity analyses to investigate the robustness of the results.	See general response
Sebastian Werner vfa	general		Sensitivity analyses are an adequate tool to **investigate and discuss the robustness of results**. This instrument should therefore play a more important role in this guidance, especially in complex data situations with different sources of evidence. In a broad sense, different sources might refer to the existence of several available clinical studies of different design/quality (randomized vs. non-randomized; one arm versus multiple arms) or direct versus indirect evidence sources or the use of different analysis methods, each with various (and sometimes untestable) assumptions. Several analyses, each based on different data sources, assumptions, or approaches, cannot completely eliminate existing uncertainties, but offer potential for comparison between the analyses. A **careful consideration and discussion of each individual analysis and the respective influencing factors is essential** for this, but also the consideration of the analyses in their entirety. This can increase the reliability of results with uncertainties of various causes. In special situations, the goal should not be the equality of all results from different analyses but **consistency**. In this context consistency	The importance of sensitivity analyses is repeatedly described in the Guideline a number of times. Additionally, we refer to the Practical Guideline D4.5.1: Applicability of Evidence.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			refers to the need to explain the differences, e.g., conservative and non-conservative sensitivity analyses have explainable differences, and can therefore provide upper or lower bounds for estimates. Therefore, the guidance should adequately consider sensitivity analysis as one potential solution to **cope with uncertainties of different sources.**	
Sebastian Werner vfa	general		The approach of testing shifted hypotheses with the requirement of **large effect estimates to draw conclusions** is mentioned in various sections of the guideline. This proposal should be removed from the guideline as it **goes far beyond methodological guidance**. The authors provide guidance on conclusions and what should be considered acceptable for decision-making. The JCA shall not draw conclusions [HTAR Article 9(1)]. **The conclusion is a value judgement** and should be left to the Member States as part of their appraisal process. Instead of providing guidance on acceptability and conclusions, the **guideline should discuss context-specific approaches to evaluate the degree of certainty based on available evidence**. Different levels of uncertainty should be discussed in specific contexts and conduction of further analysis, e.g., sensitive analysis, to evaluate robustness of the results should be promoted.	The use of shifted hypothesis testing is an accepted method to ensure formally that a treatment effect exceeds a certain threshold (e.g., in noninferiority trials). This approach is described as option.
Sebastian Werner vfa	general		The "**Practical Guideline D4.3.1 Direct and Indirect Comparison**" is referenced throughout this document. As this document is still under development and has not been released for consultation yet, question marks remain after reading the draft guideline as specifics are unclear.	Thank you – the timelines for this deliverable have also been published and will also be subject to public consultation.
Sebastian Werner vfa	general		It needs to be clarified how in case of **availability of both direct and indirect comparisons**, the indirect comparisons be considered (other than for assessing the consistency)?	We do think this is covered within the box Key Points III.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Sebastian Werner vfa	general		It should be added more explicitly that an **advanced statistical expert** should be included for assessing any indirect comparisons.	The guideline already states specific statistical expertise is advised. Moreover, the competence that are required for conducting JCA are the purpose of the D5.3.1 guideline "
Sebastian Werner vfa	general		**Subgroup analyses** are mentioned on several occasions. It would be good to have a separate chapter on this topic, in which the corresponding requirements are summarized.	We refer to the Practical Guideline D4.5.1: Applicability of Evidence, which also contains sections on subgroup analyses.
Sebastian Werner vfa	general		The **Estimand framework** had been established "to align the clinical study objective with the study design, endpoint, and analysis to improve study planning and the interpretation of analysis". This framework is also important for evaluating relative effectiveness for HTA purpose, either by direct or indirect comparisons. The guidance should adequately consider that **estimands may vary by stakeholders** and also that different intercurrent event strategies are informative and supportive for HTA decision-making.	This is out of the scope of this guideline. We refer to the Practical Guideline D4.2.1: Scoping Process.
Sebastian Werner vfa	general		The guideline distinguishes between use of individual patient level data (IPD) and aggregated data with a preference for IPD (see Section 3.5). The guidance should also consider practices to reconstruct time-to event data at the individual level. This approach should be considered given the aim of the guidance to provide the best estimate of relative effectiveness with the least uncertainty – based on the best available evidence.	We included references to methods to reconstruct IPD for time-to-event and binary endpoints.
Sebastian Werner vfa	general		There should be ongoing **scientific discourse** between assessors, academia, and industry to discuss and consider latest developments in the methodology. Ideally, **platforms** are established (independent of specific guideline review processes) **to promote exchange**.	Thank you – no response required as any future discourse as described will be under the governance of the CG.
Sebastian Werner vfa	general		The guideline should be open for innovative methods established after this guideline comes into effect. A corresponding **review process for an update of the guidelines should be implemented** to ensure that the guideline reflects the current state-	Review period and process will be identified under the CG

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			of-the art.	
M. Ermisch – GKV-Spitzenverband	General		The „Methodological Guideline on Direct and Indirect Comparisons" aims at covering a very important and broad field of assessment methods. It spans from core topics of types of evidence considered to all kinds of quantitative evidence synthesis methods. Throughout the text, numerous comments and qualifications are made regarding the reliability of evidence (in particular on randomised vs. non-randomised trials) and the appropriateness of different approaches to (in-)direct evidence synthesis. Potential limitations and the need for additional assumptions in newer approaches are mentioned. This is particularly the case in the conclusion section, which overall seems adequate. However, we are concerned that the guideline draft in its entirety does not sufficiently distinguish between which evidence base is necessary and which evidence base is preferable. It seems to focus on possible approaches for cases, when the evidence base is limited. One cannot stress enough that the best approach for the assessment of effectiveness is the availability of a limited number of well-conducted and relevant (PICO-scheme) randomised controlled trials. This alone can safeguard against the introduction of well-known problems and uncertainties, in particular (dis-)similarity and heterogeneity, which are already present in meta-analyses of direct comparisons. These problems increase with the use of indirect comparisons and again with the use of more advanced approaches (like "population adjusted methods", section 5.3). In our opinion, the latter must currently be characterised as highly experimental. In the foreseeable future, their usefulness and reliability are unclear. This should be addressed more explicitly in the guidance.	We agree with this reviewer; however, we acknowledge that MSs have different thresholds of acceptability around evidence types. We have tried to strike a balance between the view expressed in this feedback and that of others advocating for a more flexible approach.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Unfortunately, the draft seems to present a conflicting perspective with its emphasis on "networks of evidence" as an (all-)encompassing concept. Its reference to "disconnected networks", which might not be networks at all, and the inclusion of studies of inferior validity such as non-randomized, non-comparative trials is questionable. We recommend distinguishing more clearly methods for evidence synthesis from methods for selecting what kind of studies to include in the synthesis. In particular, the primary consideration of inferior study designs should be avoided in the context of evidence synthesis. It should be clarified that including studies of weaker designs cannot be justified just because more advanced synthesis methods might technically allow this. The draft's presentation of networks conveys the impression that the best approach (i.e. analysing a limited number of relevant and well-conducted studies) was only one "special case" and not the preferred one. In our opinion, the draft should be restructured to avoid this. It should be clarified that methods for direct or indirect meta-analysis of diverse studies are foremost supplementary tools to compensate for the lack of adequate direct comparative trials. However, direct comparisons are generally and in almost all cases preferred to indirect comparisons. Indirect comparisons introduce additional uncertainties and additional assumptions need to hold that (per definitionem) cannot be empirically tested at the time of assessment. In our view, this generally outweighs the controversial claim that indirect comparisons might increase reliability of estimates even in the presence of direct evidence that is sometimes made regarding mixed treatment comparisons. There is a need for careful evaluation regarding the risk of an (artefact) of increase in statistical power at the expense of introduction bias.	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Of foremost importance is the ability to reliably answer the question, whether a quantitative estimation of a relevant treatment effect is possible at all. As detailed in some of the subsequent remarks, situations may occur in which even indirect comparisons will not be available or will not allow for an estimate. Estimates that are not sufficiently reliable should not be presented. A more clear-cut guidance and discussion of this question seems warranted, as it might also influence what kind of comparisons are being considered in scoping: If the feeling is established that indirect comparisons will always enable conclusions on all kinds of hypothetical comparisons, this might induce the proliferation of comparative questions. While we recognize that it cannot be the primary responsibility of a methodological guidance to frame the questions, some consideration should be given to the likely consequence of such guidance. Generally the "possible" (in terms of methods) should be more clearly distinguished from the "preferable" (and, in consequence, the "possible" in terms of sufficient evidence for assessments). We expect that, in future, stakeholders will look for guidelines not only in terms of a scientific discussion of possibilities but also as a guidance of what studies and analyses to conduct or require. It would be unfortunate, if inadvertently impressions of an "anything goes" fashion were created – and, thus, a thinking was induced in stakeholders that, e.g., direct comparison trials might be waived in favour of a variety of indirect comparison methods to take their place. We trust that these aspects will also be considered in D4.3.1 "To produce a practical guideline on how to deal in practice with indirect comparisons.", which is upcoming according to the D4.3 COMPARATORS AND COMPARISONS Project Plan, Version 1.0, 03/12/2021.	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			In addition, we miss an assessment of the question that is highly relevant in HTA: the integration of effects on several endpoints (at least morbidity, safety and often mortality). Difficulties already present in direct comparisons (e.g. lack of reporting of endpoints, different operationalisations of endpoints, different affection by effect modification) may become aggravated if indirect comparisons are added, eventually resulting in very different networks with different issues for different endpoints, which might lead to false balancing of effects and hence false conclusions.	
Marjorie Morrison, Lymphoma Coalition	General		**<u>Matching Adjusted Indirect Comparison (MAIC)</u>** Studies indicate that the matching adjusted indirect comparison (MAIC) method is fundamentally useful in weighting to aggregate characteristics with assumptions however, it appears to have practical limitations. While MAIC is methodologically sound, the use of MAIC measurements with respect to prognostic characteristics/patient characteristics may be of concern. This will require that the use of the MAIC method statistically take into consideration and/or adjust for key variables such as: (a) differences in patient characteristics (b) where data or information, such as individual patient data and/or patient characteristics, are missing or insufficient (c) where variables exist with respect to average baseline characteristics of the patient population. Fundamentally the application of the MAIC method is intended to significantly reduce the risk of biases between study designs and provide robust comparative evidence, however, the **potential and/or perceived limitations of MAIC may warrant additional measures, assessment and/or analysis to ensure the risk of biases are removed.**	Thank you – we believe the challenges have been included in Section 5 and are further expanded in the corresponding Practical Guideline.
Marjorie Morrison,	General		**<u>Network Meta-Analysis (or Indirect Comparisons)</u>**	Thank you for your observations. This is out of the scope of this guideline.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Lymphoma Coalition			Historically, evidence derived from randomised controlled trials (RCTs) is widely considered to be the gold standard of safety and efficacy evidence as RCTs produce data with low risk of bias. In the context of the Joint Clinical Assessment (JCA), other types of evidence, such as observational studies, are discouraged as the results are perceived to be "highly uncertain" and do not provide a "meaningful estimate of the relative treatment effectiveness." Hierarchical and less weighted network meta-analysis (NMA) models incorporate data from the historical gold standard of RCTs however, patients in RCTs are selected and therefore, may not reflect or mirror the target real-world population in routine clinical practice, as patients may present with influential issues, including co-morbidities. Additionally, data insufficiencies may adversely affect the assessment and decision-making processes, real-world evidence (clinical) and real-world data may provide comparative and/or valuable clinical data to support decision-making. In consideration of direct and indirect comparisons in health technology assessment (HTA), there are regulatory and HTA models in regions of the globe that acknowledge other sources of data in assessment processes, primarily to further enrich RCT evidence or direct comparisons. For instance: 1. The Canadian Agency for Drugs and Technologies in Health (CADTH) has established a dedicated Steering Committee responsible for supporting the development of a "pan-Canadian strategic framework for the use of real-world evidence in regularly and reimbursement decision-making for drug products" in Canada. https://www.cadth.ca/real-world-evidence-decision-making 2. The United States Food and Drug Administration (FDA) also recognises real-world evidence and real-world data as sources of information, with studies helping to address	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			knowledge evidence gaps while informing oncology formulary or regulatory decision-making. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence In accordance with the assumptions of exchangeability, homogeneity, similarity, and consistency, real-world evidence and real-world data sources are positioned to provide complimentary clinical and patient data, in particular where evidence and data insufficiencies are most problematic and/or where there is a lack of robust data specific to innovative therapeutics or medicines. Therefore, in support of regulatory decision-making and robust health technology assessment processes - and as demonstrated in other regions of the world - **a real-world evidence / real-world data framework (for evaluation and to clearly define infrastructure, processes, and more) may further support regulatory decision-making while addressing data insufficiencies.**	
Matias Olsen, EUCOPE	General		One of the main reasons why meta-analyses or indirect comparisons have a difficult standing in Germany is due to the heterogeneity of the study populations of the different studies and how the matching worked out. Bearing such context in mind, the discussion should rather be more about which values for the heterogeneity/matching are seen as acceptable and less about which methodology to apply.	Thank you. As this is a methodological guideline we have focussed on the methods.
Edwards Lifesciences	General		Edwards Lifesciences believes that any proposed assessment scenario should consider **a holistic definition of patient benefit** going beyond direct measures of evidences and outcomes relevant to patients. **Improvement of hospital organizational efficiencies should be also factored in the comparison parameters for medical devices, where relevant.** Any consideration for the evidence needs:	This comment is outside of the scope of this deliverable.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			- should ensure the **predictability** of the process and its outcomes for the innovative technology developers - encourage innovative approaches in conducting assessment and generating the evidence and **move to a lifecycle approach** in the evidence generation (*Tarricone et al (2020). Lifecycle evidence requirements for high risk implantable medical devices: a European perspective. Expert Review of Medical Devices, 17:10, 993-1006, DOI: 10.1080/17434440.2020.1825074*). - The **adaptability of the methodologies** could consider the following 3 dimensions: o **nature of the technology** (i.e. implantable MD, CDx, MDx, digital technologies….) o **nature of the disease** (i.e. Cardiovascular diseases, diabetes, oncology…) o **evidence needs**: identify the minimum sufficient dataset (not necessarily limited only to the regulatory evidence requirement) in line with the product lifecycle through early interaction with both regulatory (expert panel and notified bodies) and HTA bodies (i.e. in early trial design; in continuous RWE/RWD collection) The above should be facilitated through **early dialogue together with the technology developers** to define the evidence needs, the timelines and the PICO criteria.	
Edwards Lifesciences	General		This methodological guideline is highly academic and well-constructed – but do not reflect the needs for the (early) assessment of a medical device within the HTA Regulation context. What's the likelihood for a MD to have multiple RCTs at this stage? What's the purpose to discuss and review meta-analyses, NMAs, using direct and indirect methods….? The document should illustrate more the limitations of RCTs and emphasize the various rationales why indirect comparison and RWD are important and could inform and estimate early the potential efficacy of a new medical device vs. its comparator (usually standard of care).	Thank you for your comment. Please refer to the introductory commentary which will provide some clarification on aspects of your comment. We believe comments in relation to innovation are outside the scope of this deliverable.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Non-RCTs will always have some limitations but it could still inform the treatment effect of a specific medical device to various stakeholders. This guideline sounds very negative (i.e. pp 141-142) and alternative adaptive trials seem set for failure – whereas it should be the focus of this document considering the early stage assessment in the process. The incremental innovation (lifecycle management) should also be clearly considered and highlighted as a good rationale of using indirect comparison – to enable a quantification of the treatment effect gains that bring the corresponding medical device innovation.	
Mihai Rotaru - EFPIA	general		**Practical guidance for conducting evidence synthesis** The summary explains that the methodological guideline (D4.3.2) has been developed for EU assessors for JCAs of health technologies. EFPIA believes a comprehensive and up-to-date overview has been provided on the assumptions, strengths, and weaknesses of existing and emerging methods of indirect comparison. It is evident that there are multiple approaches for conducting indirect comparisons, all of which are underpinned by different assumptions and data requirements. HTDs would welcome further practical guidance on the context specific factors (i.e., data availability) that should be considered when selecting methods(s) for direct and indirect comparison. Currently, it is unclear how the research question, the context of evaluation and the associated available evidence should be considered in methods selection. To avoid confusion, EFPIA believes the methodological guideline would also benefit from being restructured based on the data availability of the research question, as opposed to the type of method. With this approach, the methodological guideline would therefore outline different data availability situations and indicate the suitable methods per situation. EFPIA requests that additional guidance or examples are provided to	Thank you for your comment. We don't believe that data availability should be a guiding principle as suggested for determining methodological rigour and therefore we don't propose to make changes based on this comment.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			aid HTDs for the selection of appropriate methods given the data availability of the specific appraisal. This guidance will provide greater clarity on the factors the assessors will take into consideration to assess the appropriateness of the methods and assumptions the HTDs have used in their indirect comparisons. Furthermore, clarity is needed on how and when assessors will determine that a given indirect comparison would be inappropriate and/or unfeasible for the HTD to conduct. EFPIA would like the proposed methods to reflect the current state of science and evidence-based medicine. As such, the methods should be open to new challenges arising from evolving medical science and the development of new therapeutic approaches, such as cell and gene therapies and their accompanying trial evidence. It should also be recognised that indirect comparisons may be used more often in the EU JCA than MS HTAs due to the potential for multiple PICOs (e.g., alternative comparators). EFPIA believes that in order to establish a JCA that meets the requirements of the EU MS, a pragmatic deliberative process which is based on the available evidence to support MS HTA decision-making is needed. As such, the primary objective should remain the consideration of the available and not exclusively the theoretically best-possible evidence.	
Mihai Rotaru - EFPIA	general		**Multiple PICOs and necessity for indirect comparisons** The proposed scoping process (D4.2) of consolidating a PICO based on a survey from EU27 MS will likely result in multiple comparators, and potentially subgroups requested for the JCA. With this approach, conducting indirect treatment comparisons will become a critical and integral part of a JCA. The requirements for an acceptable indirect comparison as set out in this draft guidance (D4.3.2) will be stringent and unfeasible for many health technologies, particularly when recognising those included in Phase I and II JCA process	Thank you for your comment. There are a number of sections providing guidance on the types of evidence that may be available including Section 6 which provides extensive guidance on non-randomised evidence. Therefore, we don't believe that there is additional information to add at this time. The development of new methods and any required guideline updates may be

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			(orphan diseases, oncology and ATMPs), given the greater potential for EU marketing authorisation on the basis of assessment with non-RCT evidence (i.e., single arm trials) and dynamic treatment pathways associated with scientific innovation. EFPIA is concerned that when combining the potential multiple PICOs and the stringent thresholds and proposals in this methodological guideline, will prevent HTDs and assessors from conducting a JCA on the basis of the most robust comparative effectiveness assessment possible with the available data. As such, not providing benefit to MS as outlined in the EU HTA regulation, '*HTA is able to contribute to the promotion of innovation, which offers the best outcomes for patients and society as a whole, and is an important tool for ensuring proper application and use of health technologies (L 458/1, 3)'.*[1] Furthermore, the strong bias towards RCT evidence and subjective view of the inapplicability of nonrandomised evidence may lead to an EU-wide access hurdle for innovative medicines in areas of high unmet need. EFPIA believes that the methodological guideline for direct and indirect comparisons should allow for flexibility in the selection the appropriate method(s) for the relevant research question. <u>References:</u> 1. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	considered under the future CG.
Mihai Rotaru - EFPIA	general		**Population-adjusted methods: Shifted null hypothesis** The methodological guideline (D4.3.2) refers to the approach of requiring a large treatment effect and the use of the shifted null hypothesis to address certain levels of uncertainty in decision-making. The approach proposes acceptable thresholds of effect and	The use of shifted hypothesis testing is an accepted method to ensure formally that a treatment effect exceeds a certain threshold (e.g., in noninferiority trials). This approach is described as an option.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			uncertainty are imposed for the results of population-adjusted indirect comparisons to inform the JCA for the purposes of decision-making. EFPIA requests the removal of this approach from the methodological guideline for the following reasons. Firstly, drawing conclusions on the relative effect sizes using this approach goes beyond the scope of the EUnetHTA21 methodological guideline as it provides instructions to assessors on the level of uncertainty that will be acceptable for pan-EU decision making. The extent of uncertainty deemed acceptable for a specific decision problem implicitly represents a value judgement, and as such, remains the responsibility of each EU MS to consider as part of its national HTA process. This is stated in the EU HTA regulation, '*It is necessary therefore that Union action is limited to those aspects of HTA that relate to the joint clinical assessment of a health technology and to ensure in particular that there are no value judgements in joint clinical assessments in order to respect the responsibilities of Member States pursuant to Article 168(7) TFEU.*' [1] Secondly, this approach (testing of shifted hypothesis) to interpret and draw conclusions on the relative effect measures is not representative of international guidelines for indirect comparison. From our research, IQWiG is the only EU HTA agency to use this test.[2] Furthermore, existing guidance on indirect treatment comparisons from leading international organisations for HTA such as ISPOR and Cochrane do not reference the 'testing of shifted hypothesis' method.[3-6] As it is not an internationally recommended approach by neither other EU HTA bodies nor academic societies, but instead a representation of a single MS preference, it is not appropriate to include this approach in the EU guideline. In addition, there is no standard rule on which these thresholds should be based, and what is acceptable can largely vary depending on disease area and/or outcomes. According to the EU HTA Regulation, individual MS	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			are permitted to conduct complimentary analyses and we suggest that an individual MS methods preferences, such as the 'testing of shifted hypothesis', is better suited to local complementary analyses versus a JCA which should take a pan-EU perspective.[1] EFPIA requests that this approach is removed and the guideline instead, discusses context-specific approaches based on available evidence. It would be more valuable for the methodological guideline to present a number of clear recommendations when assessing the validity of population adjustment approaches (requiring a multi-faceted approach), to describe different levels of uncertainty in specific contexts and recommend further analyses which can be conducted to explore the sensitivity of the results due to the uncertainty. Reference to this approach is also made in the following sections of the methodological guideline and should be removed accordingly: ▪ Page 5, line 122-126 ▪ Page 5, line 138-140 ▪ Page 24, line 710-711 ▪ Page 26, Key points IV; bullet 4. <u>References:</u> 1. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance). 2. Institute for Quality and Efficiency in Health Care (IQWiG). General Methods. Version 6.1 of 24 January 2022. Available at: https://www.iqwig.de/en/about-us/methods/methods-paper/. 3. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, Boersma C, Thompson D, Larholt KM, Diaz M, Barrett A. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Health. 2011 Jun;14(4):429-37. doi: 10.1016/j.jval.2011.01.011. PMID: 21669367. 4. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health. 2011 Jun;14(4):417-28. doi: 10.1016/j.jval.2011.04.002. PMID: 21669366. 5. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014 Mar;17(2):157-73. doi: 10.1016/j.jval.2014.01.004. Erratum in: Value Health. 2016 Jan;19(1):121. PMID: 24636374. 6. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.	
Mihai Rotaru - EFPIA	general		**Data availability settings and appropriate methods for generation of comparative evidence** This methodological guideline (D4.3.2) currently acknowledges the presence of different data availability settings; however, it is currently structured by method of direct or indirect comparison. For example, in Section 5 for population-adjusted comparisons and Section 6 for non-randomised comparisons, it is unclear whether the text refers to connected or disconnected evidence networks, or anchored versus unanchored comparisons. To avoid confusion for HTDs and assessors, we believe the methodological guideline would benefit from being restructured based on the data availability of the research question. As such, the	Thank you for your comment. First of all, we want to highlight that the main audience for methodological guideline are the assessors and co-assessors. We acknowledge the suggested change in the structure. After long discussions about the general structure of the guideline, we are convinced that the current structure is understandable and accessible for assessors and co-assessors. However, we moved section "5.3.4 Population-adjusted methods in comparisons of single-arm trials" to the

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			guideline would outline different HTD data availability situations and provide the suitable comparison methods per situation for the purposes of the JCA. EFPIA recommends that the document is restructured to provide guidance on the following core data availability situations: **<u>Direct evidence available:</u>** 1. Availability of multiple RCTs for the intervention under evaluation versus the relevant comparator(s). Currently covered under direct methods for evidence synthesis (Section 4). **<u>No direct evidence available with all of the relevant comparator(s):</u>** 2. Availability of RCTs with common comparator(s). Methods based on comparison of relative treatment effects for the treatment under evaluation versus the relevant comparator(s): a. Where valid in case of no bias due to treatment effect modifiers (Sections 5.1 and 5.2) b. Potential bias due to imbalances in treatment effect modifiers can be adjusted for by using anchored MAIC/STC (Section 5.3) 3. No availability of RCTs with common comparator (included single-arm trials) a. Access to IPD for the intervention of interest and aggregate level data for comparator(s) of interest (Section 5.3.4) b. Access to IPD for all relevant comparator(s) (Section 6). The relevant approaches are propensity score based (IPW-matching) and/or covariate or multivariable adjustment. Since these methodological approaches have their own underlying assumptions, EFPIA believes the validity of these assumptions should be presented and substantiated by the HTD, and subsequently assessed within the scope of the JCA on a case-by-case basis.	chapter 6.1, because comparisons in disconnected networks are better placed there, which is also reflected in your suggested structure.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			EFPIA recommends the description of these data availability situations be included in this methodological guideline, as it will enable greater clarity for HTDs regarding the acceptability of methodological approaches and guidance. We believe this is particularly important when recognising those included in Phase I and II JCA process (orphan diseases, oncology and ATMPs), given the greater potential for EU marketing authorisation on the basis of assessment with non-RCT evidence (i.e., single arm trials) and dynamic treatment pathways associated with scientific innovation. EFPIA understands that the proposed practical guideline (D4.3.1) may provide greater technical guidance on the practical methodology selection (e.g., MAIC versus STC, or IPW-matching versus covariate adjusting). This should be clarified in the methods guidance.	
Mihai Rotaru - EFPIA	general		**Indirect comparisons: Single-arm, disconnected and non-randomised evidence** EFPIA is concerned by the strong tone used throughout the document against evidence synthesis methods for indirect comparisons in data availability situations such as single-arm, disconnected and non-randomised evidence. For example, Section 5.3.4 indicates issues with using MAIC/STC as population-adjusted methods for comparisons of single-arm trials and these being 'highly problematic'. The same wording is used to describe approaches using observational data requiring IPD for the comparator. Furthermore, Section 6 (Conclusions) mentions, in reference to using methods for single arm/disconnected studies etc., "…the certainty of the results provided by these techniques remains controversial." EFPIA is concerned that unanchored approaches for indirect comparisons with single-arm trials could be dismissed in a JCA, or not fully considered, based on the wording in the draft guidance.	Please see the general statement at the beginning of this document around the intention of the document. We believe that this will clarify the issues raised within this comment.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			There is published methodological guidance on the use of these approaches (e.g. NICE DSU TSD 17[1] and 18[2]) and many published examples of using these approaches in the literature. Although, limitations and interpretation of the results need to be considered carefully, and the approaches used should be tailored to the available evidence in each case, these are still considered valid approaches for many HTA agencies and international HTA societies. EFPIA requests a more balanced discussion on this topic be included given the proposed Phase I and II JCA process, given the greater potential for EU marketing authorisation on the basis of assessment with non-RCT evidence (i.e., single arm trials) and dynamic treatment pathways associated with scientific innovation. As such, EFPIA wishes to highlight this in the context of the EU HTA regulation, which states that, '*Methodologies for performing joint clinical assessments and joint scientific consultations should be adapted to include specificities of new health technologies for which some data may not be readily available. This may be the case for, inter alia, orphan medicinal products, vaccines and advanced therapy medicinal products (L 458/5; Section 24).*' [3] <u>References:</u> 1. Faria, R., Hernandez Alava, M., Manca, A., Wailoo, A.J. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015. Available from http://www.nicedsu.org.uk. 2. Phillippo, D.M., Ades, A.E., Dias, S., Palmer, S., Abrams, K.R., Welton, N.J. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submission to NICE. 2016. Available from http://www.nicedsu.org.uk. 3. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			relevance).	
Mihai Rotaru - EFPIA	general		**Subgroup analyses** EFPIA would like to highlight that the methodological guideline (D4.3.2) does not currently discuss breaking randomisation in the context of an RCT when assessing subgroups. If an RCT is stratified based on a requested subgroup definition, then it may be deemed appropriate to follow randomised methods. However, if the requested subgroup breaks randomisation, then non-randomised methods may be considered. EFPIA recommends that subgroup analyses should be considered and further discussed in Section 5 of the methodological guideline. EFPIA would also welcome further reference to and guidance on the use of subgroup analysis in direct and indirect comparisons. Subgroup analyses of treatment effect modifiers that have not been pre-specified nor have credible, biological plausibility should be actively discouraged as it avoids the risk of data dredging and false conclusions being drawn.[1,2] References: 1. European Medicines Agency. Guideline on the investigation of subgroups in confirmatory clinical trials. 2019 (available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-subgroups-confirmatory-clinical-trials_en.pdf) 2. Jansen et al. Indirect Treatment Comparison/Network Meta-analysis Study Questionnaire to Assess Relevance and Credibility to Inform Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report. Value in Health 2014; 17: 157-173.	We refer to the Practical Guideline D4.5.1: Applicability of Evidence, which also contains sections on subgroup analyses.
Mihai Rotaru - EFPIA	general		**New and emerging methods direct and indirect comparison**	Thank you – text has been added to indicate that the guideline will be

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Current wording: "The objective of this document is to describe the methods currently available for direct and indirect treatment comparisons regarding their underlying assumptions, strengths and weaknesses." Suggested rewording: "The objective of this document is to describe the methods currently available for direct and indirect treatment comparisons regarding their underlying assumptions, strengths and weaknesses. **The document will be updated according to methodological developments in the literature to ensure the most current methods are covered."** [note: **bold** denotes suggested inclusion] Rationale: To be in line with international standards and support the EU HTA Regulation calls for state of the art medical science, it is important that the guidance is regularly updated to contain the most recent methods for conducting direct and indirect comparisons. The objective of the methodological guideline is to describe currently available methods, therefore it is essential that the document is regularly updated, and remains flexible to enable the use of novel evidence synthesis methods. For example, Section 5.2.4 of the methodological guideline relating to NMA of time-to-event data, states that *"Other emerging methods for time-varying hazard ratios described in the literature may also be considered"* (p24, lines 692-694). EFPIA welcomes such openness and believe it should apply in general, not only for NMA of time-to event data. Methods of comparisons are constantly evolving and in order to ensure the methodological guidance is up-to-date and reflective of the latest methods, we recommend adding a corresponding statement in the summary and conclusion sections.	reviewed and updated where necessary.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			EFPIA wishes to highlight this recommendation in the context of the EU HTA regulation, which states that, "*The Coordination Group should ensure that the scientific joint work as well as the procedures and methodology for the preparation of joint clinical assessment reports and joint scientific consultation outcome documents guarantee the highest quality, are prepared in a timely manner, and reflect the state of the art of medical science at the time of their preparation' (L 458/5; Section 23).'* [1] <u>References:</u> 1. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Mihai Rotaru - EFPIA	general		**HTD and assessor meeting for proposed evidence synthesis** As per EFPIA's response to EUnetHTA D4.2 Scoping Process practical guideline, we believe it is important to ensure the implementation of a procedural step for consultation in the form of a scoping meeting between HTD and assessors to align on the evidence synthesis methods to be applied in the context of the evidence base for the planned JCA. This constructive dialogue would create understanding of the available evidence base, alignment on the anticipated analyses for the JCA and likely improve the overall efficiency of the JCA process for all stakeholders. In addition to a joint scoping meeting, EFPIA also recommends continual dialogue with key stakeholders, including the HTD, throughout the JCA process.	Thank you but it is not the purpose of this guideline to describe the operational processes of a JCA.
Mihai Rotaru - EFPIA	general		**Methodological guideline: recommendations versus prescriptive requirements** The methodological guideline (D4.3.2) currently transitions from	See general response

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			providing guidance in certain sections, to stipulating prescriptive requirements in other sections. For example, in the Summary, the tone and wording of the guidance is overtly prescriptive, *"If any of these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness (p4, lines 84-86)."* It is the understanding of EFPIA that the objective of the guideline is to provide guidance, rather than a set of minimum standards required to conduct direct and indirect comparisons. For example, the EU HTA regulation states, "*Joint clinical assessment shall result in a joint clinical assessment report that shall be accompanied by a summary report. Those reports shall not contain any value judgement or conclusions on the overall clinical added value of the assessed health technology and shall be limited to a description of the scientific analysis (L 458/3, Article 9.1).*"[1] As such, EFPIA is concerned in the context of the EU HTA regulation, whereby the application of minimum standards may be deemed to implicitly incorporate a value judgement as opposed to representing a scientific assessment based on the available evidence. EFPIA has also indicated this feedback applies to the following sections in the document: ▪ Page 4, line 84-86 ▪ Page 5, line 134-136 ▪ Page 12, line 320-322. Therefore, EFPIA suggests the document is revised throughout to reflect the objective of the methodological guideline, i.e. to guide HTDs and assessors to select the most appropriate methods and justification of assumptions in the JCA submission dossier to enable the successful implementation of the EU HTA regulation. <u>References:</u> 1. Regulation (EU) 2021/2282 of the European Parliament and of	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Mihai Rotaru - EFPIA	general		**Practical reporting of ITC results** EFPIA would welcome further guidance on the evidence synthesis reporting standards, future planned updates to the methodological guidelines and presentation of method selection rationale to be described in the current EUnetHTA methodological guideline, or proposed inclusion in the planned practical guideline (D4.3.1) to enable clarity for HTDs.	The corresponding Practical Guideline describes the reporting requirements for the JCA report.
Mihai Rotaru - EFPIA	general		**Acceptability of JCA dossier** EFPIA recommends that greater clarity is provided in the methodological guidelines on how and when assessors will reach a decision that a given indirect comparison is inappropriate and/or unfeasible for the HTD to conduct, so that HTDs can prepare accordingly. It is also important to put in place a robust and systematic process enabling a meaningful dialogue between the HTD and assessors during an ongoing assessment (focus of consultation D7.1) so that discussions on the analysis and evidence provided can take place. Furthermore, the consequence of non-compliance from Article 10 of the Regulation needs further expansion in the scoping process and methodology, in the context of 'feasibility'. If analyses requested by MS are not feasible due to complex or incomplete evidence networks, then some evidence requested may not be possible to generate. Since the evidence requested directly derives from the Scope, this feasibility issue needs to be addressed jointly, in a meaningful scoping meeting with the HTD.	Thank you but it is not the purpose of this guideline to describe the operational processes of a JCA.
Mihai Rotaru -	general		**Population-adjusted methods for indirect comparison**	Section 5.3 is about connected

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
EFPIA			Throughout the section on population-adjusted methods for indirect comparisons (Section 5.3), it is unclear if the text refers to an anchored/connected networks or unanchored comparisons/disconnected network. EFPIA suggests that greater clarification is needed throughout the section for the benefit for HTDs and assessors.	networks. Sub-section 5.3.4 was an exception and was indeed about disconnected network. The content has been moved to Section 6 about disconnected networks for clarity.
Mihai Rotaru - EFPIA	general		**Bayesian analysis and interpretation – EUnetHTA methodological guideline (2015)** The previous EUnetHTA methodological guideline (2015) included a discussion regarding Bayesian interpretation for evidence synthesis which have not been included in this updated guideline. This relates to the following extraction [page 16, paragraph 4]; *"In contrast to the frequentist approach, the output of a Bayesian approach can be interpreted in terms of probabilities.(26) Thus Bayesian approaches facilitate the computation of useful measures such as the probability of treatment superiority, the probability that a clinically meaningful difference exists, or the probability of clinical equivalence.(31) It is possible to compute the probability that a given treatment is the best amongst those analysed and how the other treatments are ranked, which constitutes information useful clinically and to decision makers.(32) Bayesian methods also facilitate predicting the utility of conducting additional head-to-head studies.(26) The benefits to the decision maker of Bayesian outputs need to be weighed against the increased complexity of the modelling and expertise required in applying these methods."* EFPIA requests the inclusion of a similar paragraph in this methodological guideline.	We aimed at keeping the Guideline as short and easy as possible. The Guideline is not a statistical textbook (see Objective) and very general statistical and philosophical deliberations are out of the scope.
Mihai Rotaru - EFPIA	general		**Randomised versus nonrandomised evidence** This methodological guideline (D4.3.2) is currently limited in	Discussion about the ethics or feasibility of randomisation is out of scope of the guideline (and also defining such

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			specifically highlighting the benefits of RCTs, however it does not indicate that non-randomised evidence may also provide valuable information, and that the totality of available evidence should be considered for the purposes of decision-making. The section on evidence fails to acknowledge instances in which RCTs are not ethical (e.g., areas of high unmet need), feasible (e.g., rare or orphan diseases), or practical (e.g., in disease areas without an established standard of care). Although single-arm trials could be subject to bias, they are a valid source of evidence in such situations and can be supplemented with external control data. Furthermore, as new interventions are being discovered in diseases with limited treatment options, the use of real-world evidence (RWE) and observational studies to inform outcomes of historical standards of care is very valuable and should be accepted as evidence for JCAs. EFPIA recommends the methodological guideline incorporate text that explicitly acknowledges situations where single-arm trials and observational data (including RWE or registries) may be required and useful to inform treatment effect and that this be done consistently throughout the methodological guideline. Furthermore, the guideline should acknowledge that there are such situations in which RCTs cannot be conducted or are not available. The current text gives the perception that only RCTs would be accepted for evidence synthesis in a JCA. Explicit inclusion of these additional evidence types will help to ensure that assessors are not biased towards ignoring non-randomised data that can provide meaningful evidence to inform MS HTAs for the purposes of decision-making. This is also repeated in the following sections of the document: ▪ Page 6: line 151-153 ▪ Page 8: line 249.	situations could be highly debated and controversial).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
GSK	General	General	There should be a constructive dialogue implemented between the assessors and the applicants to ensure a methodological exchange before prespecifying and conducting the corresponding analyses for a specific dossier.	Thank you but it is not the purpose of this guideline to describe the operational processes of a JCA.
GSK	General	General	The guideline should be open for innovative methods established after this guideline comes into effect. A corresponding review process for an update of the guidelines should be implemented to ensure that the guideline reflects the current state-of-the art.	Review period and process will be identified under the CG
GSK	General	General	In general, all sources of available evidence should be used. The corresponding evidence level of the sources should be considered. Instead of complete rejection of existing evidence due to violation of strict criteria, the general approach should be to **use the best available evidence**. Complete exclusion of existing evidence introduces another risk of selection bias and should therefore be avoided in the context of an evidence-based approach.	Thank you for your comment. Please refer to the general comments provided which explains the purpose of this document.
GSK	General	General	The document contains several statements indicating the strong limitations of various methods – suggesting that the default is to reject these methods. Instead, the default should be to apply innovative methods in order to use the best available evidence and to acknowledge for potential limitations, e.g., by conducting various sensitivity analyses to investigate the robustness of the results.	We described the limitations of the methods as they are. Indeed, this may lead to the situation that the application of a specific method is not appropriate if the required data for that method are not available.
GSK	General	General	The "Practical Guideline D4.3.1 Direct and Indirect Comparison" is referenced throughout this document. What is the process for reviewing / finalizing the Practical Guideline?	Timelines are outlined on the EUnethHTA website. It follows the same process as for this guideline.
GSK	General	General	In case of availability of both direct and indirect comparisons, will the indirect comparisons be used at all (other than for assessing the consistency)?	We do think this is covered within the box Key Points III.
GSK	General	General	It should be added more explicitly that an advanced statistical expert should be included for any indirect comparisons.	The guideline already states specific statistical expertise is advised. Moreover, the competence that are required for conducting JCA are the purpose of the D5.3.1 guideline "
GSK	General	General	Subgroup analyses are mentioned on several occasions. It would be good to have a separate chapter on this topic, in which the	We refer to the Practical Guideline D4.5.1: Applicability of Evidence, which

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			corresponding requirements are summarized.	also contains sections on subgroup analyses.
GSK	General	General	No mention of multiple imputation for missing demographic or baseline disease data, as well as imputation of time-to-event data.	The document is not intended to cover strategies for handling missing data in clinical studies.
GSK	General	General	Given the need to understand the extent of bias that may be present in any particular analysis when interpreting the results, suggest a discussion of QBA methods be included	QBA methods are out of the scope of this Guideline. We refer to the Practical Guideline D4.6.1: Validity of Clinical Studies.
Norbert Gerbsch for IGES Institut GmbH and HealthEcon AG	general	-	The methodological guideline provides a comprehensive overview. So far it does not provide hands-on support for the methodological practitioner, which will probably be supplied by the EUnetHTA21 practical guidelines D4.3.1, D4.2.1, F4.5.1 and D4.6.1 which are currently under development.	Yes, it will.
Roche	General		The "original" EUnetHTA Methodology Guideline Comparators & Comparisons: Direct and indirect comparisons. V2.0. 2015 concluded the summary section with the sentence *"The choice of methodology is ultimately context specific and should be appropriate to the data available."* We believe such pragmatism is important also in the context of the EU HTA Regulation for several reasons. On one hand, the first products subject to JCA will be cancer drugs and ATMPs (wave 1), followed by orphan medicinal products (wave 2). For such products challenges may arise due to small population size, novel trial designs (such as basket and platform trials), or uncontrolled trials. Flexibility and pragmatism will be necessary to be able to assess such products. In addition, the EU level policy questions as defined during the scoping process will often be broader than any single member state policy question and, therefore, may lead to a more complex evidence base. Again context specific decisions on the most appropriate approach will often be needed. The proposed updated guidance seems to be lacking an explicit	We added this sentence in the last paragraph of the Summary.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			acknowledgment that both the choice of appropriate methods and of the appropriate evidence base needs to be context specific. We recommend introducing such an explicit statement to ensure the needed flexibility and the ability to cope with the broad spectrum of evidence configurations we expect to see in practice.	
Roche	General		There should be explicit openness to emerging methods. The proposed guidance states in *Section 5.2.4 NMA of time-to-event data* that *"Other emerging methods for time-varying hazard ratios described in the literature may also be considered"* (p.24, l.692-694). We welcome such openness and believe it should apply in general, not only for NMA of time-to event data. Therefore we recommend adding a corresponding statement to the conclusion section.	Text has been added to indicate that the guideline will be reviewed and updated where necessary.
Richard Birnie, Lumanity	General		Throughout the guidance, it is concluded that estimation of relative effects in a disconnected network using current available population-adjustment methods (MAIC and/or STC) are not appropriate and do not provide a meaningful estimate of the relative effect. In many cases, single-arm trials are the only evidence available. For example, in rare diseases or in tumour agnostic indications where basket trials are commonly used. The use of single trials has increased in recent years and is likely to continue doing so with the continued drive toward more rapid licensing and reimbursement. It is unhelpful to decision makers to conclude that no meaningful estimate of the relative effect can be obtained in those cases where the alternative is no evidence at all. We fully agree that evidence derived from single arm trials or disconnected networks carries significant limitations. However, it would be preferable to provide more structured guidance on the interpretation and limitations of such analysis than to conclude that no evidence is available.	The document does not describe methods on the basis of available data; rather it describes the limitations that may arise when applying certain methods to certain evidence types. We do not consider the availability of data to be relevant in the determination of methodological rigour. However, we have altered the wording to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Further to this point, the guidance implies that access to full IPD for all treatments is required when analysis of non-randomised evidence is required. The developers of a health technology will rarely have access to IPD for comparator treatments if the evidence for the comparator comes from studies performed by another company. As such this requirement seems unachievable. If this is not the intended interpretation of the guidance then it would be beneficial to be more explicit in what level of IPD is required.	
Bayer	general		The proposed methods should reflect the current state of science and evidence-based medicine. Furthermore, the methods should be open to new challenges arising from the development of new therapeutic approaches such as cell and gene therapies and their accompanying trial evidence. This is especially true given that indirect comparisons may be used more often within Euro-HTAs than in national HTAs due to potentially different PICO schemes. In order to establish a European HTA that meets the demands of the member states, the maxim should apply: Pragmatism before methodological rigorism. HTA should be seen as a pragmatic deliberative process, which, based on the available evidence, should support the decisions of relevant decision-makers. Thus, its primary goal should also remain the consideration of available and not exclusively theoretically best-possible evidence.	We don't agree that pragmatism should completely replace methodological rigour and given this is a methodological guideline we consider it appropriate that methodological rigour is clearly described.
Bayer	general		Throughout the text, the term "adjusted" is used to refer to indirect comparisons that are "anchored" by a common comparator in a connected network (e.g., lines 110-114, Key Points III, Section 5.1). Nevertheless, the term "adjusted" is also used to describe "population-adjusted" indirect comparisons or covariate-adjusted analyses more generally (e.g., line 91, lines 115-140, Section 5.3). As highlighted in Section 5.3.4, "population-adjusted" indirect comparisons or covariate adjustment techniques can be applied in situations without a common comparator, allowing for the inclusion of single-arm trials. According to the first definition of "adjusted", these comparisons would be "unadjusted". Therefore, a suggestion	We have described what is meant by "adjusted" in relation to indirect comparisons. The term "adjusted" is used frequently when describing statistical techniques. We have read the whole text again and we don't believe there is further need for clarification here however.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			would be to reserve the term "adjusted" for covariate- or population-adjusted analyses and to use the term "anchored" for situations where there is a common comparator.	
Bayer	general		The text highlights that input from statisticians is advised for a "critical assessment of (…) assumptions potentially violated" (lines 143-146 and lines 882-885). While input from a statistician is crucial to evaluate whether exchangeability is plausible, it is worth noting that the exchangeability assumption is not inherently statistical. Any approach based on statistical testing will, almost invariably, be underpowered to assess treatment effect heterogeneity, particularly in individual RCTs (lines 375-376). Besides from statistical expertise, the identification of effect modifiers will likely require prior background knowledge, and substantive clinical and domain expertise. Clinical expertise is necessary to examine other threats to exchangeability: variability and dissimilarities across studies in patient inclusion criteria, interventions, outcome definitions, and study design. Therefore, clinical knowledge is as important as statistical knowledge to carefully consider the underlying assumptions of the methods.	We agree that clinical expertise will be required, and this is included and also expanded within the under development practical guideline.
Bayer	general		It is surprising not to see a dedicated section on IPD network meta-regression, given that: (1) it can account for treatment effect heterogeneity (as noted in lines 91-92) due to differences in patient characteristics; (2) it can be used to test for heterogeneity (as noted in lines 382-384); and (3) it can produce estimates in any target population. The target does not necessarily have to be that of the studies included in the meta-regression but could be defined by an external dataset, e.g., observational data, real-world data, registries, etc.	It is out of the scope of the Guideline to include dedicated sections on each special method.
James Ryan AZ		General	We welcome the update to the direct and indirect comparison guideline which now incorporates different techniques commonly used with HTA across Europe. In its current form, the guideline provides limited direction on which approaches should be considered dependent upon contextual factors, such as the available evidence base. It also runs the risk of the HTD	Thank you for your comment. After long discussions about the general structure of the guideline, we are convinced that the current structure is understandable and accessible for assessors and co-assessors, in particular in relation to the practical guideline.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			and assessors needing to apply all techniques which is scientifically inappropriate, inefficient and won't aid decision-making at a country level. We would propose that the guideline is restructured so that it guides the assessor (and the HTD) as to which group of techniques could be most appropriate given different evidence base situations.	
James Ryan AZ		General	The role of the HTA assessor and co-assessor is to estimate the most likely incremental benefit based on the available evidence at the time of the assessment, whilst reflecting on the evidence level. Indirect comparisons will be a critical and necessary part of Joint Clinical Assessment and are accepted across the majority of HTA bodies within Europe and around the world. Throughout the document, there are occasions where the tone moves from scientific discussion and guidance towards a prescriptive position that such comparisons are unlikely to be appropriate (for example, imposing stringent conditions that must be met), undermining the role of indirect comparisons and compromising the ability of the assessors to do their scientific role. The consequences of the document tone can bias the assessors and lead to JCA Reports that have limited benefit at a Member State level as they cannot provide any comparative benefit estimate based on this guideline. This would have a particular disproportionate effect on those medicines without RCT evidence or with low sample size, including rare diseases and innovative medicines accepted by the regulator as delivering in high unmet need diseases. The guideline should adopt a guidance tone rather than prescriptive tone, and account for the real, pragmatic challenges that HTA faces. Minimum standards should not be set; instead, the guideline should describe the strengths and weaknesses of each approach, how different types of evidence can help inform the analysis, practical	We thank the reviewer for the comments and for clarity we have provided some introductory wording to explain any misunderstanding as to the intended purpose of the methodological guideline. We have appropriately described the limitations when certain methods are used under certain conditions. We have deliberately not included value judgements in the guideline as per the HTA Regulation. The regulation does stipulate that assessors should describe the limitations around the evidence and with this in mind we have produced a guideline to enable this to be done.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			considerations, and direction when one (or a set of) technique(s) may be more appropriate than others. The Regulation scope states that no value judgement should be undertaken at a European HTA level, remaining the responsibility of Member States. Any statement that could be interpreted as a value judgement in the guideline, such as outlining stringent criteria that must be met, should be removed; the challenges and uncertainty should be discussed as part of the assessment based on the medicine and evidence context, and Member States left to apply their own stringent criteria at a national level.	
James Ryan AZ		General	Shifted hypothesis This has no basis in European or internationally recognised methods for direct and indirect comparisons and, to our knowledge, is only used by IQWiG. Given this, such an approach has no place in a European HTA guideline or assessment and should be removed. Furthermore, it should be interpreted as a value judgement as it imposes minimum thresholds on effect and levels of acceptable uncertainty. Acceptable levels of uncertainty and minimum effect size are value judgements and therefore a Member State responsibility, as stated in the Regulation. Any reference to this and treatment effect acceptability should be removed throughout the guideline. The guideline instead should focus on the role of scientifically appropriate, non-value judgement, sensitivity analyses to help demonstrate consistency and confidence in effect, taking account of the context of the disease and evidence base.	The use of shifted hypothesis testing is an accepted method to ensure formally that a treatment effect exceeds a certain threshold (e.g., in noninferiority trials). This approach is described as an option.
James Ryan AZ		General	It is important that the guideline remains up to date, so that state-of-the-art emerging techniques are regularly assessed and incorporated. A regular review process should be recommended.	Review period and process will be identified under the CG

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
EFSPI	general		Any evidence synthesis including direct and indirect comparisons should aim to maximise the value of the information available. We request an inclusive approach to using the best available evidence. Ideally, the best available evidence is data from randomized controlled trials. However, for example in orphan medicinal products or rare disease areas, single-arm trials, observational data, or even indirect comparisons based on single-arm trials, may be the best available evidence. Despite a preference towards randomized trials, the HTA regulation has provision for other evidence types "[…] when such studies are accessible." (L458/6, introduction, par 36). Accordingly, the draft guidance should not reject such evidence, for example by stating that a certain type of evidence is highly problematic. Instead, the draft guidance should encourage a careful assessment of the level of evidence, utilize the most appropriate method for evidence synthesis based on the evidence available, and plan for appropriate sensitivity analyses to investigate the robustness of conclusions to potential data limitations.	Thank you - given this is a methodological guideline we consider it appropriate that methodological rigour is clearly described. It is appropriate to describe the limitations associated with certain data types and the application of certain methods given certain situations. We agree that appropriate sensitivity analysis can be useful in exploring uncertainty where it arises and this is outlined in the guideline.
EFSPI	general		The methodology for indirect treatment comparisons is constantly evolving. When appropriately applied, emerging and innovative methods can help reduce uncertainty and facilitate better and faster decision making. This aligns with the objective of the JCA laid out in the HTA regulation (L458/6, introduction, par 36), namely to "[...] effectively facilitate market access and contribute to the timely availability of innovative health technologies for patients." Accordingly, We request the draft guidance to refrain from rejecting particular methodologies. Instead, it should promote an inclusive approach, maintaining an openness towards emerging methods with HTDs justifying their selection of methods relative to the evidence available. To allow the assessors to judge the appropriateness of non-standard and emerging methodology, the rationale for the methodology should be clearly explained (including relevant references to allow	Thank you – we believe that we have clearly outlined the limitations where they apply as is required by a methodological guideline.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			proper understanding of the methods), and results should be accompanied by a discussion of the potential limitations, as well as sensitivity analyses to investigate the robustness of conclusions to assumptions and which can quantify/understand potential bias. When novel methodologies are used, statistical experts should be involved in assessing the appropriatness of the selection of methodologies given the evidence available.	
EFSPI	General		We request an inclusive approach to different evidence types as well as emerging methodology, facilitated by clear requirements for reporting and sensitivity analyses. Specifically, we recommend that the guidance document advises the following for reporting: 1) The assumptions behind statistical approaches or the use of a specific type of evidence should be clearly stated and evaluated qualitatively, and quantitatively, using appropriate sensitivity analyses 2) There should be a 'proportionality principle', in which increasingly non-standard evidence and statistical modelling methodologies will call for increasingly detailed argumentation and sensitivity analyses 3) Sensitivity analyses should distinguish between testable and non-testable assumptions, and it should be clearly described what assumption(s) is investigated in a given sensitivity analysis. - Sensitivity analyses of untestable assumptions should adopt a structured approach in which the result is viewed as a function of an assumption which is then varied in a systematic way (for example, Phillippo et al. 2017, DOI: 10.1111/rssa.12341; Liublinska and Rubin 2014 DOI: 10.1002/sim.6197; Ding and VanderWeele 2016 (DOI: 10.1097/EDE.0000000000000457). This will allow individual member states to judge the range of results/conclusions that are possible for a range of assumptions that they find meaningful with respect to the member state	Reporting requirements are provided in the upcoming Practical Guideline: Direct and Indirect Comparisons (deliverable D4.3.1). However, these reporting requirements refer more to the HTDs and what should be included in the submission dossier. We point to the Guideline D5.1: "Submission Dossier Guidance".

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			setting.	
EFSPI	General		Taken together, draft guidances D4.2 and D4.3.2 raise principled concerns about the risk of over-interrogation of data. The Joint Clinical Assessment will simultaneously assess potentially many different PICOs across different evidence bases, producing a potentially very large number of statistical analyses. Each of these statistical analyses will be visible to all Member States. There is a real risk of generating confusion among HTA bodies, prescribers, and patients, for example if conclusions across such two related PICOs are conflicting due to random chance. It is acknowledged that the intention is to increase transparency. However, this must be accompanied by increased accessibility by member states of the interpretability of the results. To increase accessibility (interpretability), it is critical that assessors are mindful of potential multiplicity issues in their assessment which may not have been considered when the clinical trial(s) were designed. To ensure that assessors consider potential multiplicity issues in their assessments, we recommend that the following is included in the draft guidance (for example, it could be added on line 222). Recommended new text: "With multiple PICOs and accompanying evidence syntheses that were not considered as part of the design of clinical trial(s), statistical multiplicity issues is a concern i.e. as the number of statistical inferences increase, so does the chance of obtaining findings that are simply due to random chance. This limitation should be carefully considered when identifying PICOs of interest. In particular, any additional significance hypothesis testing not accounted for in pre-specified multiplicity testing strategies should be considered descriptive and interpreted with caution"	The definition of PICOs as well as the issue of multiplicity are covered by other Guidelines (D4.2.1 and D4.5.1).
EFSPI	General		The draft guidance does not provide any considerations around the	We refer to the PICO framework, which

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			role of estimands in connection with direct or indirect treatment comparisons (ICH E9 (R1)Addendum. Estimands and sensitivity analysis in clinical trials. EMA/CHMP/ICH/436221/2017). We acknowledge that estimands are yet not widely incorporated into HTA. Conversely, the use of estimands in a regulatory setting will only increase in the years to come, and estimands will be a key part of both the protocol and statistical analysis plans of ongoing and future randomized trials for registration purposes. To appropriately guide assessors faced with the notion of estimands during assessments, it is critical to provide basic guidance on how to account for estimands in a direct and indirect treatment comparisons. This is in line with ICH E9 (R1), which states "A naïve comparison between data sources, or integration of data from multiple trials without consideration and specification of the estimand that is addressed in each data presentation or statistical analysis, could be misleading." Recommended new text (e.g. in Section 3.1): "Different randomized trials may rely on different estimands (ICH E9 (R1)Addendum. Estimands and sensitivity analysis in clinical trials. EMA/CHMP/ICH/436221/2017), reflecting different clinical questions of interest. Evidence synthesis should take into account the type of estimands reported and should aim to combine estimands that address the same clinical question. There should be a thorough discussion in case a mixture of estimands are available"	is very similar to the estimand. The statistical data analysis of single studies is not the scope of the Guideline (see Objective).
EFSPI	General		We request that the guidance clarifies what is understood by subgroup analyses since this is a frequent matter of discussion in practice. Specifically, we encourage that the draft guidance clarifies the distinction between proper subgroup analysis – potentially pre-specified in the trial – where subgroups are derived from characteristics obtained prior to randomization, and post-randomization subgroup analysis, derived from characteristics obtained after randomization such as effects among responders or adherers (Yusuf et al. 1991, JAMA, 266(1):93-98). Analysis of	We refer to the Practical Guideline D4.5.1: Applicability of Evidence, which also contains sections on subgroup analyses.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			proper subgroups enjoys the benefit of randomisation whereas the analysis of post-randomisation subgroups do not and will generally require alternative methods such as those used in observational data analysis to properly account for identifying subgroups based on factors observed post-randomisation. Recommended new text: "Subgroups should be based on characteristics that are measured prior to randomization. These subgroups will have the benefit of randomization, and subgroup treatment effects can be estimated directly in an unbiased manner. This is different from subgroups based on variables defined post-randomisation, for example analysis of treatment effects in responders. Post-randomisation subgroups generally require observational methods for analysis since subgroup membership may be correlated with treatment."	
EFSPI	General		The draft guidance does not provide recommendations regarding reporting of indirect treatment comparisons along the lines of, for example CADTH (2015), Guidance document on reporting indirect comparisons. It is unclear if guidance will be part of the dossier template. Actionable guidelines for the reporting will simplify both the assessment, the subsequent use of the assessment, and the comparison across assessments. If and when guidelines are developed, we recommend to consult with stakeholders from scientific and professional societies such as ISPOR, EFSPI, PSI and special interest groups with appropriate expertise.	Thank you. There is additional work on Practical Guidelines to complement these Methodological guidelines. The consultation process for involvement of stakeholders under EUnetHTA21 is outlined on the EUnetHTA website.
EFSPI	General		The guidance mentions pre-specification in several places, as a way to address the uncertainty that arises for example in relation to choice of effect modifiers and confounders. We agree that pre-specification is a useful tool to mitigate the risk of cherry picking adjustment strategies that improve results. However, it is not fully clear yet how pre-specification can be operationalized by the HTD in the context of JCA. To be able to accommodate the timelines, the	We anticipate that the provision of JSC will aid HTDs to consider the pre-specification that may be required for JCA at an early stage.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			HTD will need to plan statistical analysis to address expected scope long before the actual scope is communicated. Without information on the actual scope, any pre-specification risks becoming a laundry list of all possible scenarios. This kind of pre-specification will not contribute towards the scientific integrity of the JCA. To ensure that pre-specification can be done in a meaningful way, we encourage an increased HTD involvement in the JCA process and in particular during the scoping process.	
Laurent Petit, Leem	General	I Introduction, objective and scope	The descriptive accumulation of methods without any prioritization based on the gradation of the evidence that each method can provide is not likely to simplify, harmonize and accelerate the European access process. At the very least, the methods selected in this list should be among those recommended by the main international HTA organizations such as ISPOR or Cochrane. Finally, in order for the JCA to be usable at the national level, the number of methods used for indirect comparisons within the same file must be limited and discussed and validated beforehand during the scoping meeting between the evaluators and the company. In order to ensure that the objective of establishing a common base of scientific evaluation that can be used for value assessment at national level is met, it is essential that the quality of the evidence is described in the joint report.	Thank you. It is not the purpose of this guideline to describe the operational processes of a JCA.
Kjetil G. Brurberg, Norwegian Institute of Public Health	General		The document is in general well written, and we choose to provide one comment that applies to the document in general. It is stated under objectives that the document is not meant to cover methods for evidence synthesis of diagnostic accuracy studies. We realize there is a need to limit the scope, but we would also stress that this is a limitation. Given the development of new drugs based on the use of companion diagnostics the world of diagnostic accuracy and effectiveness trials are brought closer together. There would be a	Thank you. Further consideration may be given under the governance of the CG.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			need for guidelines on how to deal with companion diagnostics (needs, availability and accuracy) in HTA about drugs that depend on the use of companion diagnostics. Hopefully this will be covered in other deliveries.	
Paolo Morgese - ARM	General comment	General comment	This guideline is on "Direct and Indirect Comparisons" and not on "Comparators and Comparisons" the guideline on "Comparators and comparisons" 4.3.1 is planned for August 2022	This is a typographical error; the titles of the Methodological Guideline (D4.3.2) and the forthcoming Practical Guideline (D4.3.1) are the same: "Direct and Indirect Comparisons".
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		The summary explains that the methodological guideline (D4.3.2) has been developed for EU assessors for JCAs of health technologies. Takeda believes a comprehensive and up-to-date overview has been provided on the assumptions, strengths, and weaknesses of existing and emerging methods of indirect comparison. It is evident that there are multiple approaches for conducting indirect comparisons, all of which are underpinned by different assumptions and data requirements. We would welcome further practical guidance on the context specific factors (i.e., data availability) that should be considered when selecting methods(s) for direct and indirect comparison. Currently, it is unclear how the research question, the context of evaluation and the associated available evidence should be considered in methods selection. To establish a JCA that meets the requirements of the EU MS, Takeda supports a pragmatic deliberative process which is based on the available evidence to support MS HTA decision-making is needed. As such, the primary objective should remain the consideration of the available and not exclusively the theoretically best-possible evidence. Additional guidance would be useful to aid HTDs for the selection of appropriate methods given the data availability of the specific appraisal. This guidance will provide greater clarity on the factors the assessors will take into consideration to assess the appropriateness	Thank you for your comment. We don't believe that data availability should be a guiding principle as suggested for determining methodological rigour and therefore we don't propose to make changes based on this comment.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			of the methods and assumptions the HTDs have used in their indirect comparisons. Furthermore, clarity is needed on how and when assessors will determine that a given indirect comparison would be inappropriate and/or unfeasible for the HTD to conduct. Finally, we request future JCA assessors be open to new methods arising from evolving medical science and the development of new therapeutic approaches, such as cell and gene therapies and their accompanying trial evidence. It should also be recognised that indirect comparisons may be used more often in the EU JCA than MS HTAs due to the potential for multiple PICOs (e.g., alternative comparators).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		The proposed scoping process (D4.2) of consolidating a PICO based on a survey from EU27 MS will likely result in multiple comparators, and potentially subgroups requested for the JCA. With this approach, conducting indirect treatment comparisons will become a critical and integral part of a JCA. The requirements for an acceptable indirect comparison, as set out in this draft guidance (D4.3.2), are stringent and may be unfeasible for many health technologies, those included in Phase I and II JCA process (orphan diseases, oncology and ATMPs) where EU marketing authorisation is more commonly granted on the basis of assessment with non-RCT evidence (i.e., single arm trials). These field also have dynamic treatment pathways associated with scientific innovation and therefore multiple likely comparators across the EU – also often requiring ITCs to demonstrate comparative effectiveness. Takeda is concerned that when combining the potential multiple PICOs and the stringent thresholds and proposals in this methodological guideline, will create a situation that prevents HTDs and assessors from conducting a JCA on with the available data. As such, the EU JCA process may not provide benefit to MS as outlined in the EU HTA regulation, '*HTA is able to contribute to the*	Thank you for your comment. There are a number of sections providing guidance on the types of evidence that may be available including section 6 which provides extensive guidance on non-randomised evidence. Therefore, we don't believe that there is additional information to add at this time. The development of new methods and any required guideline updates may be considered under the future CG.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			promotion of innovation, which offers the best outcomes for patients and society as a whole, and is an important tool for ensuring proper application and use of health technologies (L 458/1, 3)'.[1] Furthermore, the strong bias towards RCT evidence and subjective view of the inapplicability of nonrandomised evidence may lead to an EU-wide access hurdle for innovative medicines in areas of high unmet need. Takeda believes that the methodological guideline for direct and indirect comparisons should allow for flexibility in the selection the appropriate method(s) based on the available evidence for the relevant research question and requests the guideline tone be revisited to this flexibility. <u>References:</u> 3. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		The methodological guideline (D4.3.2) refers to the approach of requiring a large treatment effect and the use of the shifted null hypothesis as an approach to set and evaluate the acceptable levels of uncertainty in decision-making. The approach proposes thresholds of effect and uncertainty for the results of population-adjusted indirect comparisons which are acceptable, or have to be met, in order for the analysis to inform the JCA for the purposes of decision-making. Takeda requests the removal of this approach from the methodological guideline for the following reasons. Firstly, drawing conclusions on the relative effect sizes using this approach goes beyond the scope of the EUnetHTA21 methodological guideline as it provides instructions to assessors on the level of uncertainty that will be acceptable for pan-EU decision making. The	The use of shifted hypothesis testing is an accepted method to ensure formally that a treatment effect exceeds a certain threshold (e.g., in noninferiority trials). This approach is described as an option.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			extent of uncertainty deemed acceptable for a specific decision problem implicitly represents a value judgement, and as such, remains the responsibility of each EU MS to consider as part of its national HTA process. The EU HTA Regulation (HTAR) states that no value judgement should take place at the joint clinician assessments. In specific, the HTAR states: '*It is necessary therefore that Union action is limited to those aspects of HTA that relate to the joint clinical assessment of a health technology and to ensure in particular that there are no value judgements in joint clinical assessments in order to respect the responsibilities of Member States pursuant to Article 168(7) TFEU.*' [1] Secondly, this approach (testing of shifted hypothesis) to interpret and draw conclusions on the relative effect measures is not a universally used or referenced approach for indirect comparison. From our research, IQWiG is the only EU HTA agency to use this test.[2] Furthermore, existing guidance on indirect treatment comparisons from leading international organisations for HTA such as ISPOR and Cochrane do not reference the 'testing of shifted hypothesis' method.[3-6] As it is not an internationally recommended approach by neither other EU HTA bodies nor academic societies, but instead a representation of a single MS preference, it is not appropriate to include this approach in the EU guideline. In addition, there is no standard rule on which these thresholds should be based, and what is acceptable can largely vary depending on disease area and/or outcomes. According to the EU HTA Regulation, individual MS are permitted to conduct complimentary analyses and we suggest that an individual MS methods preference, such as the 'testing of shifted hypothesis', is better suited to local complementary analyses versus a JCA which should take a pan-EU perspective.[1] Takeda requests that this approach be removed and the guideline instead discusses context-specific approaches based on available evidence. It would be more valuable for the methodological guideline to present a number of clear recommendations when assessing the	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			validity of population adjustment approaches (requiring a multi-faceted approach), to describe different levels of uncertainty in specific contexts and recommend further analyses which can be conducted to explore the sensitivity of the results due to the uncertainty. Reference to this approach is also made in the following sections of the methodological guideline and should be removed accordingly: ▪ Page 5, line 122-126 ▪ Page 5, line 138-140 ▪ Page 24, line 710-711 ▪ Page 26, Key points IV; bullet 4. <u>References:</u> 7. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance). 8. Institute for Quality and Efficiency in Health Care (IQWiG). General Methods. Version 6.1 of 24 January 2022. Available at: https://www.iqwig.de/en/about-us/methods/methods-paper/. 9. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, Boersma C, Thompson D, Larholt KM, Diaz M, Barrett A. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health. 2011 Jun;14(4):429-37. doi: 10.1016/j.jval.2011.01.011. PMID: 21669367. 10. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health. 2011 Jun;14(4):417-28. doi: 10.1016/j.jval.2011.04.002. PMID: 21669366.	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			11. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014 Mar;17(2):157-73. doi: 10.1016/j.jval.2014.01.004. Erratum in: Value Health. 2016 Jan;19(1):121. PMID: 24636374. ▪ Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		This methodological guideline (D4.3.2) is currently structured by methods of direct or indirect comparison. As a result, the current structure may cause confusion to the applicability and reference of the text based on the data availability context. For example, in Section 5.3 for population-adjusted comparisons and Section 6 for non-randomised comparisons, it is unclear whether the text refers to connected or disconnected evidence networks or anchored versus unanchored comparisons. To avoid confusion for HTDs and assessors, we believe the methodological guideline would benefit from being restructured based on the data availability of the research question. As such, the guideline would outline different HTD data availability situations and provide the suitable comparison methods per situation for the purposes of the JCA. We recommend that the document is restructured to provide guidance on the following core data availability situations: **Direct evidence available:** 4. Availability of multiple randomised controlled trials (RCTs) for the intervention under evaluation versus the relevant comparator(s). Currently covered under direct methods for	Thank you for your comment. First of all, we want to highlight that the main audience for methodological guideline are the assessors and co-assessors. We acknowledge the suggested change in the structure. After long discussions about the general structure of the guideline, we are convinced that the current structure is understandable and accessible for assessors and co-assessors. However, we moved section "5.3.4 Population-adjusted methods in comparisons of single-arm trials" to the chapter 6.1, because comparisons in disconnected networks are better placed there, which is also reflected in your suggested structure.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			evidence synthesis (Section 4). **<u>No direct evidence available with all of the relevant comparator(s):</u>** 5. **Anchored analysis**: Availability of comparative evidence (randomised or not) with common comparator(s). Methods based on comparison of relative treatment effects for the treatment under evaluation versus the relevant comparator(s): a. Where valid in case of no heterogeneity or imbalance in treatment effect between studies (Sections 5.1 and 5.2) b. Potential bias due to imbalances in treatment effect modifiers which can be adjusted for by using anchored MAIC/STC or multi-level network meta-regression (Section 5.3) 6. **Unanchored analysis**: No availability of evidence with common comparator (including single-arm trials and disconnected networks) a. Access to IPD for the intervention of interest and aggregate level data for comparator(s) of interest (Section 5.3.4) b. Access to IPD for all relevant comparator(s) (Section 6). The relevant approaches are propensity score based (IPW-matching) and/or covariate or multivariable adjustment. Since these methodological approaches have their own underlying assumptions, the validity of these assumptions should be presented and substantiated by the HTD, and subsequently assessed within the scope of the JCA on a case-by-case basis. Takeda recommends the description of these data availability situations be included in this methodological guideline, as it will enable greater clarity for HTDs regarding the acceptability of methodological approaches and guidance. We believe this is particularly important when recognising those included in Phase I and II JCA process (orphan diseases, oncology and ATMPs), given the greater potential for EU marketing authorisation on the basis of assessment with non-RCT evidence (i.e., single arm trials) and	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			dynamic treatment pathways associated with scientific innovation. We anticipate that the proposed practical guideline (D4.3.1) may provide greater technical guidance on the practical methodology selection (e.g., MAIC versus STC, or IPW-matching versus covariate adjusting). This should be clarified in the methods guidance.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	General Examples: 26 27-29	Examples: Section 5.3.4 Section 6	Takeda is concerned by the tone and strong language used throughout the document against evidence synthesis methods for indirect comparisons in data availability situations such as single arm, disconnected and non-randomised evidence. For example, Section 5.3.4 indicates issues with using MAIC/STC as population-adjusted methods for comparisons of single-arm trials and these being 'highly problematic'. The same wording is used to describe approaches using observational data requiring IPD for the comparator. Furthermore, Section 6 (Conclusions) mentions, in reference to using methods for single arm/disconnected studies etc., "…the certainty of the results provided by these techniques remains controversial." Takeda is concerned that unanchored approaches for indirect comparisons with single-arm trials could be dismissed in a JCA, or not fully considered, based on the wording in the draft guidance. There is published methodological guidance on the use of these approaches (e.g. NICE DSU TSD 17[1] and 18[2], ZIN[3], TLV[4], PBAC[5]) and many published examples of using these approaches in the literature. An analysis of different HTA body approached to the use of ITC and their preferred methods found that, '*ITC is generally accepted as a technique that allows demonstration of noninferiority to a comparator provided the chosen methodology* and underlying assumptions are clear and *justified*.'[6] Although, limitations and interpretation of the results need to be considered carefully, and the approaches used should be tailored to the available evidence in each	Please see the general statement at the beginning of this document around the intention of the document. We believe that this will clarify the issues raised within this comment.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			case, these are still considered valid approaches for many HTA agencies and international HTA societies. We request a more balanced discussion on this topic be included given the proposed Phase I and II JCA process, given the greater potential for EU marketing authorisation on the basis of assessment with non-RCT evidence (i.e., single arm trials) and dynamic treatment pathways associated with scientific innovation. As such, Takeda wishes to highlight this in the context of the EU HTA regulation, which states that, *'Methodologies for performing joint clinical assessments and joint scientific consultations should be adapted to include specificities of new health technologies for which some data may not be readily available. This may be the case for, inter alia, orphan medicinal products, vaccines and advanced therapy medicinal products (L 458/5; Section 24).'* [7] <u>References:</u> 4. Faria, R., Hernandez Alava, M., Manca, A., Wailoo, A.J. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015. Available from http://www.nicedsu.org.uk. 5. Phillippo, D.M., Ades, A.E., Dias, S., Palmer, S., Abrams, K.R., Welton, N.J. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submission to NICE. 2016. Available from http://www.nicedsu.org.uk. 6. Zorginstituut Nederland, ZIN (2017). www.zorginstituutnederland.nl/over-ons/publicaties/rapport/2016/09/09/procedure-beoordeling-extramurale-geneesmiddelen 7. Dental and Pharmaceutical Benefits Agency, TLV (2017). www.tlv.se/lakemedel/ansok-om-pris-eller-subvention/ 8. Pharmaceutical Benefits Advisory Committee, PBAC (2017). https://pbac.pbs.gov.au/ 9. Ischa, J.E., Spoors, J. (2018) Analysis of indirect treatment	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			comparisons in national health technology assessments and requirements for industry submissions. Journal of Comparative Effectiveness Research. 1. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		Current wording: "The objective of this document is to describe the methods currently available for direct and indirect treatment comparisons regarding their underlying assumptions, strengths and weaknesses." Suggested rewording: "The objective of this document is to describe the methods currently available for direct and indirect treatment comparisons regarding their underlying assumptions, strengths and weaknesses. **The document will be updated according to methodological developments in the literature to ensure the most current methods are covered."** [note: **bold** denotes suggested inclusion] Rationale: To be in line with international standards and support the EU HTA Regulation calls for state of the art medical science, it is important that the guidance is regularly updated to contain the most recent methods for conducting direct and indirect comparisons. The objective of the methodological guideline is to describe currently available methods and therefore it is essential that the document is regularly updated, and remains flexible to enable the use of novel evidence synthesis methods. For example, Section 5.2.4 of the methodological guideline relating to NMA of time-to-event data, states that *"Other emerging methods for time-varying hazard ratios*	Thank you – text has been added to indicate that the guideline will be reviewed and updated where necessary.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			described in the literature may also be considered" (p24, lines 692-694). Takeda welcomes such openness and believe it should apply in general, not only for NMA of time-to event data. Methods of comparisons are constantly evolving and in order to ensure the methodological guidance is up-to-date and reflective of the latest methods, we recommend adding a corresponding statement in the summary and conclusion sections. This recommendation is supportive of the EU HTA regulation, which states that, "*The Coordination Group should ensure that the scientific joint work as well as the procedures and methodology for the preparation of joint clinical assessment reports and joint scientific consultation outcome documents guarantee the highest quality, are prepared in a timely manner, and reflect the state of the art of medical science at the time of their preparation' (L 458/5; Section 23).'* [1] <u>References:</u> 2. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		Aligned to the Takeda and in support of the EFPIA response to EUnetHTA D4.2 Scoping Process practical guideline, we believe it is important to ensure the implementation of a procedural step for consultation in the form of a scoping meeting between HTD and assessors to align on the EU PICO and therefore the appropriate evidence synthesis methods to be applied in the context of the evidence base for the planned JCA. This constructive dialogue would create understanding of the available evidence base, alignment on the anticipated analyses for	Thank you but it is not the purpose of this guideline to describe the operational processes of a JCA.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			the JCA and likely improve the overall efficiency of the JCA process for all stakeholders. In addition to a joint scoping meeting, we also recommend continual dialogue with key stakeholders, including the HTD, throughout the JCA process.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		The methodological guideline (D4.3.2) currently transitions from providing guidance in certain sections, to stipulating prescriptive requirements in other sections. For example, in the Summary, the tone and wording of the guidance is overtly prescriptive, *"If any of these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness (p4, lines 84-86)."* It is our understanding that the objective of the guideline is to provide guidance, rather than a set of minimum standards required to conduct direct and indirect comparisons. For example, the EU HTA regulation states, "*Joint clinical assessment shall result in a joint clinical assessment report that shall be accompanied by a summary report. Those reports shall not contain any value judgement or conclusions on the overall clinical added value of the assessed health technology and shall be limited to a description of the scientific analysis (L 458/3, Article 9.1).*"[1] As such, Takeda is concerned in the context of the EU HTA regulation, whereby the application of minimum standards may be deemed to implicitly incorporate a value judgement as opposed to representing a scientific assessment based on the available evidence. This feedback applies to the following sections in the document: ▪ Page 4, line 84-86 ▪ Page 5, line 134-136 ▪ Page 12, line 320-322. Therefore, Takeda respectfully requests the document be revised throughout to reflect the objective of the methodological guideline,	See general response

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			i.e. to guide HTDs and assessors to select the most appropriate methods and justification of assumptions in the JCA submission dossier to enable the successful implementation of the EU HTA regulation. References: ▪ Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		Further guidance is requested on the evidence synthesis reporting standards, future planned updates to the methodological guidelines and presentation of method selection rationale to be described in the current EUnetHTA methodological guideline, or proposed inclusion in the planned practical guideline (D4.3.1) to enable clarity for HTDs.	Future planned updates and additional guidance will be under the governance of the Coordination Group and its subgroups.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	general		This methodological guideline (D4.3.2) is currently limited in specifically in that is focuses on the benefits of RCTs, but does not indicate that non-randomised evidence may also provide valuable information, and that the totality of available evidence should be considered for the purposes of decision-making. The section on evidence fails to acknowledge instances in which RCTs are not ethical (e.g., areas of high unmet need), feasible (e.g., rare or orphan diseases), or practical (e.g., in disease areas without an established standard of care). Although single-arm trials could be subject to bias, they are a valid source of evidence in such situations and can be supplemented with external control data. Furthermore, as new interventions are being discovered in diseases with limited treatment options, the use of real-world evidence (RWE) and observational studies to inform outcomes of historical standards of care is very valuable and should be accepted as evidence for JCAs.	Discussion about the ethics or feasibility of randomisation is out of scope of the guideline (and also defining such situations could be highly debated and controversial).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Takeda recommends the methodological guideline incorporate text that explicitly acknowledges situations where single-arm trials and observational data (including RWE or registries) may be required and useful to inform treatment effect and that this be done consistently throughout the methodological guideline. Furthermore, the guideline should acknowledge that there are such situations in which RCTs cannot be conducted or are not available. The current text gives the perception that only RCTs would be accepted for evidence synthesis in a JCA. Explicit inclusion of these additional evidence types will help to ensure that assessors are not biased towards ignoring non-randomised data that can provide meaningful evidence to inform MS HTAs for the purposes of decision-making. This is also repeated in the following sections of the document: ▪ Page 6: line 151-153 ▪ Page 8: line 249.	
Takeda	general		Throughout the document, and particularly in Section 6, 'confounders' and 'prognostic factors' are used interchangeably. It is unclear if in EUnetHTA21's interpretation 'confounders' also includes prognostic factors. For avoidance of confusion, Takeda requests the language be clarified.	The document has been checked for consistency.
MTE	General		- **MedTech Europe calls for the consideration of innovative approaches** adapted to the characteristics of medical technologies and solutions to perform comparative analysis. This to have **more flexibility in pragmatic and proportionate evidence generation over the lifecycle of the technology.** - This entails adapted data generation both in early trial design as well as in the continuous RWE/RWD collection, in line with the evolving nature of these innovations and the	We don't agree that pragmatism should replace methodological rigour and given this is a methodological guideline we consider it appropriate that methodological rigour is clearly described

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			acceptance of innovative approaches of comparative analysis (including indirect comparisons) - Evidence generation should take into account the availability of RWE: this to appreciate the current actual effectiveness of care delivery and the remaining unmet needs. With the aid of RWE, taking into account contextual factors (i.e., the users' proficiency or 'learning curve', training, interpretation, multiple indications, pace of the technologies' innovation cycle, the adaptation of the care pathways, etc.) evolving comparative effectiveness data and evidence will be available.	
Matias Olsen, EUCOPE		309-312	Even though there is no gold standard method regarding disconnected networks, the use of evidence from disconnected networks needs to remain a possibility in certain situations like rare diseases, advanced therapy medicinal products (ATMPs) and paediatric populations. This would then be in line with the conception of evidence based medicine in general, that states: "The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research." and "Evidence based medicine is not restricted to randomised trials and meta-analyses. It involves tracking down the best external evidence with which to answer our clinical questions." as well as "And if no randomised trial has been carried out for our patient's predicament, we must follow the trail to the next best external evidence and work from there." (See Sackett 10.1136/bmj.312.7023.71). Add: "The use of such evidence in JCA is highly problematic because the EU regulation requires comparative results on the basis of adequate comparisons (PICO framework) [26]. **If there are no RCTs**	See general response

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			available, the best available evidence should be provided by the HTD to address the PICO research question. Any provided evidence should be considered by the assessors acknowledging different levels of evidence.".	
Bayer	general	Section 3.1	The consistency assumption for NMA should be discussed also in the assumptions section (not only in Section 5 or later) as one additional homogeneity aspect that additionally arises in this context. It should be described that a Cochran-type chi-squared Q statistic can be calculated for the overall heterogeneity in the network and can be decomposed into components accounting for heterogeneity among studies comparing the same treatments and inconsistency among studies comparing different treatments. See e.g. • *Krahn et al.: A graphical tool for locating inconsistency in network meta-analyses. BMCMedical Research Methodology 2013 13:35. http://www.biomedcentral.com/1471-2288/13/35* • *Schwarzer et al. Meta-Analysis with R. Basel, Switzerland: Springer; 2015.*	After a long discussion with various changes of the general structure, we decided to present the main assumptions applicable to all methods in section 3.1, whereas assumptions that are specific to a subset of methods are described in the corresponding section.
GSK	13-14	356-367	Although the term similarity in this context has been defined as being restricted to effect modifiers, it's not clear that later uses of the term are similarly restrictive (e.g. "If dissimilarities between studies in study design and patient characteristics are observed at a level that is considered substantial, …."). Suggest clear statement that the term "similar" has this specific meaning throughout the document.	Thank you for this comment. However, it would be very difficult to have a 100% consistency of terminology and think this is not strictly required. Although we have e.g., the "balance assumption" in Section 6, of course the term "balanced design" has a different meaning. We tried to use a consistent terminology as far as possible and hope that the respective meaning is clear from the context.
Bayer	15-16	Section 3.2	This section discusses potential sources of bias for classical meta-analysis however does not discuss specific topics for the more complex NMA or MAIC approaches.	No, this section discusses general sources of bias for meta-analysis and other forms of evidence synthesis. Specific topics for NMA are discussed in Section 5.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Bayer	16-17	Section 3.4	This section discusses rather unspecific a few aspects of frequentist versus Bayesian approaches but does not address the specific context of direct and indirect comparisons.	Yes, this short section discusses a few general issues for frequentist and Bayesian methods. The specific issues for direct and indirect comparisons are described later in the respective sections.
Sebastian Werner vfa	4	70-72 / Summary	Single as well as multiple sources require appropriate methods. Please delete "in the presence of multiple sources of evidence" in the sentence "To assess the relative efficacy or effectiveness of a new intervention compared to one or more existing interventions (the comparators, e.g., the current standard treatment) in the presence of multiple sources of evidence, appropriate methods for evidence synthesis should be used.".	The guideline is about evidence synthesis. The methods required for single sources of evidence are described under other guidelines.
Sebastian Werner vfa	4	84-85 / Summary 609-611 / 5 Indirect comparison Key points III	It is emphasized that if certain assumptions i.e., sufficient similarity, sufficient homogeneity, sufficient consistency for comparisons are violated, then "the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness." We believe that it depends on the context and potential impact of the underlying decision problem. Thus, still in case of deviations estimates of treatment effectiveness could be supportive and address uncertainties. Thus, we suggest replacing the sentence "If any these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness." by "Violations of this assumptions will impact the results of the corresponding evidence synthesis and the conclusiveness of the estimate of treatment effectiveness."	We believe the proposed statement does not provide meaningful information to assessors. Nonetheless, we do agree that the acceptability of an NMA is on spectrum rather than binary. Thus, we have amended the text to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".
Sebastian Werner vfa	4 6	79–81/Summary 167 – 169 / I	The guideline is highly relevant also for other stakeholder like pharmaceutical companies. Suggestion for rewording: "The guideline is aimed at assessors and other stakeholders like pharmaceutical companies…"	We agree that the Guideline is relevant for other stakeholders. However, it is aimed at the assessors.
Mihai Rotaru - EFPIA	4	84 / Summary	**Summary: Assumption violation and meaningful estimates**	See general response

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Current wording: "If any of these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness." Suggested rewording: "~~If any of these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness.~~" [note: ~~strikethrough~~ denotes proposed deletion] Rationale: The term 'meaningful' relates to interpretation of treatment effect and therefore represents a value judgement, which is the responsibility of each MS and out of scope for the JCA. Also, as the criteria for violation of assumptions are not given, this statement leaves too much room for interpretation and should be removed.	
Mihai Rotaru - EFPIA	4	86 / Summary Repeated 17; 504	**Context: Heterogeneity** Current wording: "If the heterogeneity is considered too strong to justify an overall evidence synthesis but the heterogeneity can be explained, appropriate evidence synthesis should be performed using the corresponding group of trials or subgroups of patients." Suggested rewording: If the **between-trial** heterogeneity is considered too strong **(for illustration, I^2 value > 75%)** to justify an overall evidence synthesis but the heterogeneity can be explained, appropriate evidence synthesis should be performed using the corresponding group of trials or subgroups of patients. Caution is needed in interpreting the results of subgroup analyses, particularly if the subgroups have not been pre-specified. Data availability is also an important consideration. Outcomes reporting for relevant subgroup	Thanks for the comment. We do not think prescribing a I^2 threshold is relevant. Regarding subgroup analyses, we have added a reference to a specific EUnetHTA 21 Practical Guideline dealing with these concerns.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			across all trials may vary which also limits the usefulness of this approach." [note: **bold** denotes proposed inclusion] <u>Rationale:</u> There are two types of heterogeneity to consider when conducting indirect comparisons 1) between trial heterogeneity 2) within trial heterogeneity. The current wording in the methods guide does not differentiate between the two types. EFPIA believes the wording "too strong" is subjective, providing details of more objective tests of between-trial heterogeneity would help to ensure consistency between EU assessors. For illustration, the Cochrane v5 handbook uses I^2 thresholds of 50% to 90% to represent substantial heterogeneity and 75% to 100% to represent considerable heterogeneity. While we are not suggesting this test statistic should be used as "a rule", it could be as a guide by assessors and HTDs. Subgroup analyses are considered observational in nature and have the same limitations as observation studies including possible bias due to observed or unobserved imbalances in prognostic factors and treatment effect modifiers. Undertaking multiple post-hoc analyses to explain heterogeneity may raise concerns of data dredging.[1] Researchers have also raised concerns over the usefulness of subgroup analysis, and meta-regression to improve trial similarity for adjusted indirect comparison. They highlight that the number of trials included in an adjusted indirect comparison is usually small and it is uncertain whether all important study level variables will be reported.[2] <u>References:</u> 1. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook. 2. Song F, Loke Y K, Walsh T, Glenny A, Eastwood A J, Altman D G et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews BMJ 2009; 338 :b1147 doi:10.1136/bmj.b1147.	
Mihai Rotaru - EFPIA	4	92 / Summary	**Editorial: Elimination of heterogeneity** Current wording: "However, while these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely." Suggested rewording: "However, ~~while~~ these methods are likely to reduce heterogeneity~~, it is unlikely that they will eliminate it completely~~." [note: ~~strikethrough~~ denotes proposed deletion] Rationale: EFPIA wishes to highlight that the methods used for this purpose have the objective of reducing heterogeneity as opposed to its elimination. As such, EFPIA suggests removal given the negative implication in the document.	Thanks. We have proposed an addition in the corresponding paragraph.
Bayer	4	73-75/ Summary	There is emphasis on RCTs being the gold standard for informing treatment effectiveness estimates due to their low risk of bias (e.g., lines 73-75, and 202-203). RCTs minimize "internal validity" bias (i.e., bias within the study sample) due to treatment arm exchangeability and no confounding on expectation. Nevertheless, there is also "external validity" bias to consider (Westreich et al. 2019). For instance, a well-conducted RCT may not necessarily be representative of the target population of policy interest, e.g., if the study sample has a different distribution of treatment effect	Thank you – external validity of any evidence is of course subject to assessment (and widely recognised in the literature). We don't believe this is specific to just RCTs and therefore we don't propose changes as suggested.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			modifiers than the target population. In that case, the efficacy estimate produced in the analysis of the RCT may not be a good estimate of real-world effectiveness. It may be worth clarifying that RCTs have the lowest risk of "internal validity" bias. To assess "external validity" bias, the "P" in PICO would have to be clearly defined. *Westreich, D., Edwards, J.K., Lesko, C.R., Cole, S.R. and Stuart, E.A., 2019. Target validity and the hierarchy of study designs. American journal of epidemiology, 188(2), pp.438-443.*	
Bayer	4	83-86/ Summary	Current: If any of these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness. Proposed: If any of these assumptions is violated, the results of the corresponding evidence synthesis are accompanied by higher uncertainty and the derived treatment effect may be questioned Rationale: A violation of an assumption does not lead automatically to a useless estimate as suggested in the above statement. Assumptions are not axioms and apodictical within hypothesis logic. Furthermore, the text implies that the assumption of consistency is required for indirect comparisons to provide meaningful estimates of treatment effectiveness. Similar statements about sufficient consistency are made in lines 162-163, 337-338, 601-605, and in Key Points III. Nevertheless, consistency is a property specific to loops of evidence. Where there are no loops of evidence, there are no direct comparisons available to contrast the indirect comparisons with. This is often the case in small networks, e.g., anchored two-study scenario with a common comparator group, where any violations of consistency cannot be assessed.	See general response
Bayer	4	90-92/ Summary	There is likely a misprint here: "If heterogeneity is caused by study characteristics rather than patient characteristics, meta-regression with adjustment for variables contributing to the heterogeneity is another option for dealing with heterogeneity." The covariate-adjusted outcome regression in a meta-regression is used to account for differences in patient characteristics, not differences in study	No, this is not a misprint. If heterogeneity is caused by patient characteristics, meta-regression (based upon aggregated data) is usually not an option to deal with this heterogeneity.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			characteristics. Heterogeneity caused by study characteristics is a more reasonable rationale for the use of random- or mixed-effects models.	
Bayer	4	92-93/ Summary	Current: However, while these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely. Proposed: these methods are likely to reduce heterogeneity. Rationale: The methods used for this purpose do not claim elimination of heterogeneity but only reduction. There is no need for a negative co-notation.	We think that it is important to say clearly that heterogeneity is still an issue after applying subgroup analyses or meta-regression.
Bayer	4	104-105/ Summary	Current: The standard frequentist approach for random effects meta-analyses is the Knapp-Hartung method in cases involving at least five studies. Proposed: Delete. Rationale: This is one of the accepted methods next to the established DerSimonian and Laird REF MA.	We disagree; the DerSimonian-Laird-method is no longer accepted; please read Section 4.1 and the corresponding references.
Bayer	4	106-108/ Summary	Current: In situations with fewer than five studies, alternative methods for evidence synthesis are frequently required, such as Bayesian approaches, a qualitative summary of the study results or the beta-binomial model. Proposed: In situations with fewer than five studies, alternative methods for evidence synthesis are frequently required, such as fixed effect models, Bayesian approaches or the beta-binomial model. Rationale: FEF MA do not account for between trial heterogeneity and therefore offer more robust estimator in cases with low number of trials than REF MA. A qualitative summary on the other hand is not an evidence synthesis method.	The sentence refers to that before, i.e., this refers to the REM situation. A qualitative summary is a form of evidence synthesis. Thus, the proposed change was not adopted.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	4	84 / Summary	Current wording: "If any of these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness." Suggested rewording: "~~If any of these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful~~	See general response

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			~~estimate of treatment effectiveness.~~" [note: ~~strikethrough~~ denotes proposed deletion] Rationale: The term 'meaningful' relates to interpretation of treatment effect and therefore represents a value judgement, which is the responsibility of each MS and out of scope for the JCA. Also, as the criteria for violation of assumptions are not given, this statement leaves too much room for interpretation and should be removed.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	4	86 / Summary Repeated 17; 504	Current wording: "If the heterogeneity is considered too strong to justify an overall evidence synthesis but the heterogeneity can be explained, appropriate evidence synthesis should be performed using the corresponding group of trials or subgroups of patients." Suggested rewording: If the **between-trial** heterogeneity is considered too strong **(for illustration, I^2 value > 75%)** to justify an overall evidence synthesis but the heterogeneity can be explained, appropriate evidence synthesis should be performed using the corresponding group of trials or subgroups of patients." [note: **bold** denotes proposed inclusion] Rationale: There are two types of heterogeneity to consider when conducting indirect comparisons 1) between trial heterogeneity 2) within trial heterogeneity. The current wording in the methods guide does not differentiate between the two types. The wording "too strong" is subjective, providing details of more objective tests of between-trial heterogeneity would help to ensure consistency between EU assessors. For illustration, the Cochrane v5	Thanks for the comment. We do not think prescribing a I2 threshold is relevant. Regarding subgroup analyses, we have added a reference to a specific EUnetHTA 21 Practical Guideline dealing with these concerns.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			handbook uses I^2 thresholds of 50% to 90% to represent substantial heterogeneity and 75% to 100% to represent considerable heterogeneity. While we are not suggesting this test statistic should be used as "a rule", it could be as a guide by assessors and HTDs. References: 2. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	4	92 / Summary	Current wording: "However, while these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely." Suggested rewording: "However, ~~while~~ these methods are likely to reduce heterogeneity~~, it is unlikely that they will eliminate it completely~~." [note: ~~strikethrough~~ denotes proposed deletion] Rationale: Takeda wishes to highlight that the methods used for this purpose have the objective of reducing heterogeneity as opposed to its elimination and therefore suggests removal given the negative implication in the document.	Thanks. We have proposed an addition in the corresponding paragraph.
James Ryan AZ	4	86-114	Few words of caution are provided on the biases and practical issues (i.e., data availability) of undertaking indirect comparison on patient subgroups. Reword: "If the heterogeneity is considered to be too strong to justify an	We don't believe all the additional wording suggested provides clarity for assessors and therefore will not include as suggested. However we will again examine the language to improve clarity.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			overall evidence synthesis but the heterogeneity can be explained, appropriate evidence syntheses should be performed using the corresponding groups of trials or subgroups of patients. This results in different effect estimates for the different subgroups. If heterogeneity is caused by study characteristics rather than patient characteristics, meta-regression with adjustment for variables contributing to the heterogeneity is another option for dealing with heterogeneity. However, while these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely" To (additions in bold, removal via cross through): "If the heterogeneity is **considered to be substantial** but the heterogeneity **in patient characteristics is known,** appropriate evidence syntheses **can** be performed using the corresponding groups of trials or subgroups of patients. This results in different effect estimates for the different subgroups. If heterogeneity is caused by study characteristics rather than patient characteristics, meta-regression with adjustment for variables contributing to the heterogeneity is another option for dealing with heterogeneity. However, while these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely. **In the case of subgroup analysis, trials may not be stratified or powered to capture treatment effects in these groups. Similarly, adjustment via meta-regression may not capture all important study characteristics**" **Same comment should be applied to similar statements throughout document**	
James Ryan AZ	4	115-140	Reword: "…may be considered provided that the network is connected and individual patient-level data are available for some of the trials included. These methods require that all effect modifiers relevant for	We don't believe all the additional wording suggested provides clarity for assessors and therefore will not include as suggested. However we will again examine the language to improve

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			adjustment are measured. However, this is often unverifiable and unattainable. Therefore, it is imperative that population-adjusted indirect comparisons are thoroughly investigated to ascertain whether these methods produce a better estimate of the treatment effect. The model and covariate selection strategies for adjustment must be prespecified and based on transparent criteria. Owing to the greater uncertainties associated with population-adjusted methods, a large treatment effect estimate is required, which can be formally achieved via testing of shifted hypotheses. This means that a conclusion can be drawn regarding an effect only if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect" To (additions in bold, removal via cross through): …may be considered provided that the network is connected and individual patient-level data are available for some of the trials included. **The robustness of these methods depending on the extent to which all important effect modifiers relevant for adjustment are measured and captured in the analysis**. ~~However, this is often unverifiable and unattainable.~~ Therefore, it is imperative that population-adjusted indirect comparisons are thoroughly investigated to ascertain whether these methods produce a ~~better~~ **more robust** estimate of the treatment effect **compared to simpler methods such as subgroup analyses**. The model and covariate selection strategies for adjustment must **ideally** be prespecified, based on transparent criteria, **and validated by clinical experts**. ~~Owing to the greater uncertainties associated with population-adjusted methods, a large treatment effect estimate is required, which can be formally achieved via testing of shifted hypotheses. This means that a conclusion can be drawn regarding an effect only if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect~~	clarity.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Same comment should be applied to similar statements throughout document	
GSK	4	102-108	Is there an expectation to follow one of the mentioned methods? Would any scientific approach with appropriate references and scientific rigour also be considered?	A stated this guideline is not proposed as an exhaustive list of methods available and of course methods develop over time. Any updates to this guideline will be under the governance of the future CG.
Paolo Morgese - ARM	4	73-76	When it comes to Advanced Therapy Medicinal Products (ATMPs), Randomised Controlled Trials (RCTs) are often not feasible. Due to the rarity of target diseases, the high unmet need and the significant clinical effect, it is often ethically not appropriate to set up controlled trials. In many cases, data filed for regulatory approvals is developed in single arm studies. This fact is widely acknowledged by the clinical, regulatory and HTA communities. Twenty ATMPs have been approved in the European Union since the EU ATMP Regulation came into force in 2009 and many HTA assessment and appraisal have been done in EU member states. Overall, there has been broad acceptance and acknowledgement of the need to develop approaches to address clinical uncertainties coming from single arm studies. These approaches include the use – among others - of: •indirect comparisons, •synthetic comparators, •clinical assumptions conditional to further evidence to be developed in real-life, •methods for extrapolation. ARM understands that RCTs are the "gold standard" for HTA of traditional medicines, especially those medicines targeting large populations and having marginal added benefit vs. standard of care. But RCTs are not the clinical standard study vehicle for investigating efficacy and safety of many ATMPs. For this reason, ARM calls EUnetHTA 21 to develop specific approaches for evaluating these non-RCT frameworks and addressing clinical uncertainties of ATMPs	Thank you. Section 6 of the guidelines provides guidance on non-randomised evidence. The GRADE methodology is outside the scope of this guideline.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			at launch, including methods for measuring uncertainty and evidence development plans for mitigating such uncertainties. It is suggested that there is a modification of the GRADE methodology to make allowances for ATMPs.	
Silke Walleser Autiero Medtronic	4	106-108	With a small number of studies Bayesian methods are preferred but sensitivity analyses should be carried out as the priors can drive inference. Also, the underlying normality assumption is often incorrect and other models can provide more meaningful (less biased) estimates. Also, prediction intervals may provide a better estimate of what is likely to be seen from a new study from the same population.	OK, but these lines belong to the summary, which is not the right place for all details.
EFSPI	4	99-101	Current wording: "Analyses based on individual patient-level data are generally preferable to aggregated data" The wording suggests that analyses based in individual patient-level data are inherently more accurate than those based on aggregate data. This is an oversimplification that may lead assessors to believe that aggregate data analyses are inherently inferior when, in fact, analysis of both will often lead to the same results and conclusions under the same modelling assumptions (Smith et al 2016, DOI: 10.1002/14651858.MR000007.pub3; Matthew and Nordström 1999 DOI: 10.1111/j.0006-341x.1999.01221.x). Recommended rewording: "Analyses based on individual patient-level data, when available, offer the potential for exploring additional and potentially more appropriate statistical analyses or subgroup analyses compared to aggregated data"	Thanks for the input. We have proposed a clarification.
Sebastian Werner vfa	4	92-93	"However, while these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely." As all sources of available evidence should be acknowledged, please add to the sentence above: "While these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely. Therefore, the magnitude, source, and potential impact of the heterogeneity on the	The added points move away from the main point. As a Summary should be as short as possible it should not be overloaded with details. The issue of heterogeneity is tackled in its own section later.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			results has to be discussed carefully and might be supported by other sources of evidence or sensitivity analyses."	
Liebenhoff, BAH	4	92-93	"However, while these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely." Since no sources should be excluded from the onset, please add the following. "Therefore, the potential impact of heterogeneity on outcomes needs to be considered in detail and could be supported by other sources of evidence or sensitivity analyses."	We do not think these details are necessary in the summary. The issue of heterogeneity is tackled in its own section later.
EFSPI	4	77-78	The draft guidance states that its' objective is "[…] to describe the methods most commonly used for direct and indirect treatment comparisons". The preferred methods for direct and indirect comparisons are constantly evolving, reflecting that this is a highly active statistical research area. To accommodate this, we recommend that 1) the guidance takes an inclusive approach to methodology and does not dismiss emerging methodologies following the EUnetHTA21 joint action period, a framework is set up for regular updating of the guideline in which stakeholders and experts in the field are consulted to ensure that the guideline reflects state of the art in direct and indirect treatment comparisons	Review period and process will be identified under the CG
EFSPI	4	92-93	Current wording: "However, while these methods are likely to reduce heterogeneity, it is unlikely that they will eliminate it completely" We suggest to remove the sentence. It is not expected to have a mean square error equal to zero. There is always the sampling error, and this cannot be adjusted by meta-regression.	Heterogeneity in the context of evidence syntheses means variability beyond sampling error. Therefore, the statement is correct.
Sebastian Werner vfa	4	74	Please change "…of treatment effectiveness and…" to "…of treatment efficacy and…". Effectiveness is more reserved for real-world study in literature.	As explained in the document, this is used as a general term. Moreover, RCTs are also studies that take place in the

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			(Line 204 explains this as for simplicity)	"real world".
EFSPI	4	76	The concept of an "adequate RCT" is mentioned in several places but remains undefined. This is a critical concept. We recommend to define as per HTAR (L458/22, art 18, par 4): "Directly comparative clinical studies which are randomised, blinded and include a control group".	Thanks for the input. We have added a reference to the upcoming practical guideline validity of clinical studies regarding what we consider an "adequate RCT".
Matias Olsen, EUCOPE	4	83	What is the difference between similarity and homogeneity for trial data? What is considered o be 'sufficient'?	Similarity and homogeneity are described in details in their section later in the document.
EFSPI	4	86	Current wording: "If any of these assumptions is violated, the results of the corresponding evidence synthesis do not provide a meaningful estimate of treatment effectiveness." The wording here is unclear and suggests that it is a case of 'either/or' whereas in practice, it is a continuum (towards increasingly biased results). Recommended rewording: "Indication that any of these assumptions are not met raises a concern that results may be biased and the extent to which assumptions are met should be discussed."	We don't agree and consider the wording to be clear. The recommended change to the text would change the meaning and would be less clear. Nevertheless, we modified the statement to "... are unlikely to provide a meaningful estimate …".
M. Ermisch – GKV-Spitzenverband	4	87	Suggestion: "strong" => "large"	We changed this.
M. Ermisch – GKV-Spitzenverband	4	88	Suggestion: "explained" => "explained in terms of characteristics of trials and trail populations" (clarification seems called for, as only explanations in such terms enable the analysis of "groups of trials or subgroups of patients")	We changed this.
Sebastian Werner vfa	4	90	Please change "heterogeneity is caused by study characteristics rather than patient characteristics" to "heterogeneity is caused by study characteristics or patient characteristics"	We revised the paragraph..
Sebastian Werner vfa	4	90	Arguably, patient characteristics such as demographic, disease severity can be better adjusted using meta-regression and study characteristics, e.g., outcome definition, rescue therapy usage	We disagree; meta-regression with aggregated data is a critical method to deal with patient characteristics.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			algorithm, are harder.	
M. Ermisch – GKV-Spitzenverband	4	93	While it is true that heterogeneity will not be eliminated completely, this is not the salient point – this being, whether sufficient reduction can be achieved. This should be reflected in the text.	The issue of heterogeneity is tackled in its own section later.
GSK	4	98	Can RWE data be used to specify the prior?	This kind of details should not be described within the Summary of the document.
Sebastian Werner vfa	4	106	Please change "situations with fewer than five studies" to "situations with fewer than five studies and only aggregated data is available". Direct comparisons often involve studies conducted by one sponsor and patient-level data can be available, this does not require more than five studies.	The 'standard' situation in meta-analyses is the case of aggregated data; therefore, this addition is unnecessary.
GSK	4 6	79 – 81 167 - 169	The guideline is highly relevant also for other stakeholder like pharmaceutical companies. Suggestion for rewording: "The guideline is aimed at assessors and other stakeholders like pharmaceutical companies…"	We agree that the Guideline is relevant for other stakeholders. However, it is aimed at the assessors.
Sebastian Werner vfa	5	110-111 / Summary Key points III	It appears that only connected evidence networks are accepted for indirect comparisons. Although they provide a higher evidence level, disconnected networks may still be informative. For instance, reference prediction and aggregate level matching may be also useful methods connecting single-arm studies or connecting disconnected RCTs for relative effectiveness evaluations for example when considering line agnostic drug labels. In this respect the sentence "If indirect comparisons are required, only adjusted indirect comparisons respecting randomisation are appropriate, which means that the evidence network must be connected. "to change to "If indirect comparisons are required, adjusted indirect comparisons respecting randomisation where evidence network has to be connected, is more appropriate than indirect comparisons based on disconnected networks."	The document provides guidance on disconnected networks as well as connected networks. The document does not propose to reject or accept different networks, rather, it provides the limitations around the application of certain methods to certain data types.
Sebastian Werner vfa	5	134-135 / Summary	The assumptions for propensity scoring approach requires sufficient positivity, sufficient overlap, and sufficient balance. We believe that it depends on the context and potential impact of the underlying	We believe the proposed statement does not provide meaningful information to assessors. Nonetheless, we do agree

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			decision problem. Thus, still in case of deviations estimates of treatment effectiveness could be supportive and address uncertainties. Thus, we suggest replacing the sentence "If any of these assumptions is not met, an adequate adjustment for confounding is not possible and the results from the corresponding analysis do not provide a meaningful estimate of treatment effectiveness" by "Violations of this assumptions will impact the results of the corresponding evidence synthesis and the conclusiveness of the estimate of treatment effectiveness"	that the acceptability of a method is on spectrum rather than binary. Thus, we have amended the text we have altered the wording to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".
Mihai Rotaru - EFPIA	5	110 / Summary	**Preference for direct versus indirect comparisons** Current wording: "If indirect comparisons are required, only adjusted indirect comparisons respecting randomisation are appropriate, which means that the evidence network has to be connected." Suggested rewording: "If indirect comparisons are required, adjusted indirect comparison respecting randomisation is ~~appropriate, which means that the evidence network has to be connected~~ **the most appropriate method, which means that the evidence network is connected. However, in cases where this is not feasible, adjustment methods are available to deal with bias due to missing randomisation."** [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] Rationale: EFPIA agrees that these methods have the least bias; however, in cases where this is not feasible, adjustment methods are available to deal with bias due to missing randomisation. These methods are explained later in the guideline, so we believe the rewording is necessary here.	Thank you for your suggestion – we will examine the wording for clarity.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Suggested rewording relates to evidence synthesis situations covering multiple comparators, and/or relevant subgroups anticipated based on the draft JCA scoping process and multiple proposed PICOs. This statement is also repeated in the following sections of the methodological guideline: ▪ Page 21, line 591-592 ▪ Page 22,key Points III, bullet 1.	
Mihai Rotaru - EFPIA	5	134 / Summary	**Editorial: Assumption violation and meaningful estimates** <u>Current wording:</u> "If any of these assumptions is not met, an adequate adjustment for confounding is not possible and the results from the corresponding analysis do not provide a meaningful estimate of treatment effectiveness." <u>Suggested rewording:</u> "If any of these assumptions is not met, ~~an adequate adjustment for confounding is not possible and the results from the corresponding analysis do not provide a meaningful estimate of treatment effectiveness~~ **the results of the corresponding evidence synthesis are accompanied by higher uncertainty and the derived treatment effect may be questioned**." [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] <u>Rationale:</u> EFPIA recommends the deletion of this statement since a violation of an assumption does not lead automatically to a meaningless	We have reworded this sentence to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			estimate.	
Mihai Rotaru - EFPIA	5	143 / Summary	**Statistical expertise** Current wording: "Input from a statistician with specific expertise in this area is advised for a critical assessment of the methodological approach used, and assumptions potentially violated and the corresponding uncertainty of the results." Suggested rewording: "Input from a statistician with specific expertise in this area is advised for a critical assessment of the methodological approach used, and assumptions potentially violated and the corresponding uncertainty of the results. **In addition, input from an experienced biostatistician who is well versed in evidence synthesis to avoid relevant information being dismissed.**" [note: **bold** denotes proposed inclusion] Rationale: EFPIA recommends the involvement of a biostatistician with experience in evidence synthesis to avoid relevant information being dismissed due to the requirement of potentially arbitrary levels of sufficient certainty.	The guideline already states specific statistical expertise is advised. Moreover, the competence that are required for conducting JCA are the purpose of the D5.3.1 guideline
Roche	5	124-126/Summary	We welcome that the document D4.3.2 provides guidance on the appropriate use of population adjustment methods. However, at various occasions the proposed guidance goes beyond a methodological guide. The sentence *"This means that a conclusion can be drawn regarding an effect only if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect."* corresponds to an instruction on the levels of uncertainty that should be acceptable for decision making. Whether	The use of shifted hypothesis testing is described as option. This sentence gives only an explanation what it means to use the approach of shifted hypothesis testing if it is used. Whether it is used is left to the Member State.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			or not a given level of uncertainty is acceptable for a specific decision with respect to a policy question should be left to Member States as part of their appraisal process. We recommend removing this as well as similar sentences in other sections, namely p.5, l.124-126; p.24, l.712-714; p.27, l.801-803.	
Roche	5	119-121/Summary	In the sentence *"Therefore, it is imperative that population-adjusted indirect comparisons are thoroughly investigated to ascertain whether these methods produce a better estimate of the treatment effect."* it may not be clear to the reader which alternative approach the method should be compared against. In addition, the term *"better"* in the above sentence is unclear. Which criteria or properties are intended here, and how can this be investigated in practice? We propose to remove the second part of the sentence, which leads to: *"Therefore, it is imperative that population-adjusted indirect comparisons are thoroughly investigated."* Alternatively, more specific guidance that is feasible in practice should be given.	Thank you for your comment. We have changed the wording for better clarity.
Roche	5	127-129/Summary	This sentence suggests that all methods for disconnected networks require IPD for all studies. We recommend updating the text to cover also methods such as unanchored STC and MAIC, which have been developed for the case of disconnected networks where a mix of IPD and AD are available - as the guidance correctly points out elsewhere.	In the document, it is described why these methods should be avoided in that context.
Bayer	5	122-140/ Summary	Unlike direct comparisons indirect comparisons are not powered a priori for their primary end point(s) at least, applying "testing of shifted hypothesis" to derive respective endpoint-specific confidence intervals implying stronger expected effects is a misleading approach	The use of shifted hypothesis testing is an accepted method to ensure formally that a treatment effect exceeds a certain threshold (e.g., in noninferiority trials).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			in regard with the reduced power of indirect comparisons. Furthermore, the testing of a shifted hypothesis implies value judgements regarding the extent of the assumed expected effect to be tested, which are not compliant with the regulation in regard of the JCA.	This approach is described as an option.
Bayer	5	134-136/ Summary	Current: If any of these assumptions is not met, an adequate adjustment for confounding is not possible and the results from the corresponding analysis do not provide a meaningful estimate of treatment effectiveness Proposed: If any of these assumptions is not met, the results of the corresponding evidence synthesis are accompanied by higher uncertainty and the derived treatment effect may be questioned Rationale: A violation of an assumption does not lead automatically to a useless estimate as suggested in the above statement. Assumptions are not axioms and apodictical within hypothesis logic.	We have reworded this sentence to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful
Tanja Podkonjak – Takeda Pharmaceuticals International AG	5	110 / Summary	<u>Current wording:</u> "If indirect comparisons are required, only adjusted indirect comparisons respecting randomisation are appropriate, which means that the evidence network has to be connected." <u>Suggested rewording:</u> "If indirect comparisons are required, **anchored** indirect comparison respecting randomisation is ~~appropriate, which means that the evidence network has to be connected~~ **the most appropriate method, which means that the evidence network is connected. However, in cases where this is not feasible, adjustment methods are available to deal with bias due to missing randomisation."** [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] <u>Rationale:</u> We agree that these methods have the least bias; however, in cases where this is not feasible, adjustment methods are available to deal	Thank you for your suggestion – we will examine the wording for clarity.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			with bias due to missing randomisation. These methods are explained later in the guideline, so we believe the rewording is necessary here. The suggested rewording relates to evidence synthesis situations covering multiple comparators, and/or relevant subgroups anticipated based on the draft JCA scoping process and multiple proposed PICOs. In addition, Takeda suggests rewording 'adjusted' to 'anchored' for clarity. This statement is also repeated in the following sections of the methodological guideline: ▪ Page 21, line 591-592 ▪ Page 22,key Points III, bullet 1.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	5	134 / Summary	Current wording: "If any of these assumptions is not met, an adequate adjustment for confounding is not possible and the results from the corresponding analysis do not provide a meaningful estimate of treatment effectiveness." Suggested rewording: "If any of these assumptions is not met, ~~an adequate adjustment for confounding is not possible and the results from the corresponding analysis do not provide a meaningful estimate of treatment effectiveness~~ **the results of the corresponding evidence synthesis are accompanied by higher uncertainty and the derived treatment effect may be questioned**." [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] Rationale:	We have reworded this sentence to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Takeda recommends the deletion of this statement since a violation of an assumption does not lead automatically to a meaningless estimate.	
EFSPI	5	122-126	Current wording: "Owing to the greater uncertainties associated with […]or below a certain threshold shifted away from the zero effect." A test of shifted hypotheses corresponds to an instruction on the level of uncertainty acceptable for decision making. This represents a value judgment and is the responsibility of member states. This approach does not provide a commonly accepted or operational way of evaluating the sensitivity to model assumptions or data limitations. It provides no insights into how results change with different assumptions, nor does it target specific model assumptions or data limitations. Instead, it requires the assessor to subjectively translate assessor-perceived model and data limitations to an anticipated bias. Tests using shifted hypotheses should thus not be a requirement. Recommended rewording: "An application of population adjustment methods should be accompanied by appropriate sensitivity analyses to evaluate the impact of data limitations and statistical modelling assumptions in a clear and targeted manner".	The use of shifted hypothesis testing is an accepted method to ensure formally that a treatment effect exceeds a certain threshold (e.g., in noninferiority trials). This approach is described as an option.
GSK	5	127-129	It may not be feasible for a sponsor to gain access to individual patient-level data when this data is being provided through third party use contracts. Can this sentence reflect a recommendation that sponsors seek access to all IPD data as this will increase the ability to fully investigate the assumptions, strengths and potential limitations in the data.	We have reflected this point in section 3.5.
Silke Walleser Autiero Medtronic	5	127-129	It is not always possible to have access to individual patient data (IPD). In the case of disconnected networks, it should not be mandatory that complete access to IPD be available before these studies can be used in the network. Other methods should also be permitted that can adequately adjust for confounders.	We have reflected this point in section 3.5.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Silke Walleser Autiero Medtronic	5	144-146	It is not just uncertainty, it is also the validity. Magnitudes and interpretations could also be wrong. Uncertainty implies variation in the point estimate.	Uncertainty is used here as a reference to certainty of results which includes internal validity.
EFSPI	5	112-114	Current wording: "[…] and the frequentist and Bayesian approaches for network meta-analysis". Pairwise indirect comparisons are a special case of network meta-analysis.	Thank you – there has been much discussion around the terms used to describe indirect comparisons – we believe we have explained this in the text adequately and therefore we do not propose to change this further as believe it may cause confusion.
EFSPI	5	127-129	Current wording: "In the case of disconnected networks (e.g., single-arm trials) and any situations with nonrandomised data, complete access to the individual patient-level data is required in order to apply methods that can adequately adjust for confounding." The draft guidance should promote an inclusive approach to using the best available evidence. While we agree that disconnected networks come with many problems and high risk of bias, we recommend all available methods relevant to the evidence available are allowed to be considered and HTDs justify the methods selected. In addition it should be noted that HTDs should make every effort to obtain IPD in order to adjust for confounding. However, there may be some situations where HTDs are not able to access IPDs, e.g. when third parties own data and it cannot be shared due to data privacy and competitiveness considerations. This may apply to oncology, when single arm trials support the development, for example in rare tumours, and where the HTD may not have access to IPD for all trials. Recommended rewording: "In the case of disconnected networks (e.g., single-arm trials) and any situations with nonrandomised data, individual patient-level data is necessary to be able to adjust for confounding. If individual patient-level data is not available, the reasons for this should be clearly described, and sensitivity analyses	We have reflected this point in Section 3.5.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			should be considered to explore the potential impact on conclusions."	
EFSPI	5	138-140	Current wording: "Owing to the greater uncertainty associated with nonrandomised data, a large treatment-effect estimate is required, which can be formally achieved via testing of shifted hypotheses." A test of shifted hypotheses corresponds to an instruction on the level of uncertainty acceptable for decision making. This represents a value judgment and should not be included in the JCA. Moreover, this approach does not provide a commonly accepted or operational way of evaluating the sensitivity to model assumptions or data limitations. It provides no insights into how results change with different assumptions, nor does it target specific model assumptions or data limitations. Instead, it requires the assessor to subjectively translate assessor-perceived model and data limitations to an anticipated bias. We recommend to refer to established epidemiological methods for evaluating the impact of unmeasured confounding in a targeted and quantitative manner, e.g. Ding and VanderWeele 2016 (DOI: 10.1097/EDE.0000000000000457) and references herein. Recommended rewording: "Owing to the greater uncertainty associated with nonrandomised data, robustness is key and appropriate sensitivity analyses should be conducted to assess the impact of unmeasured confounding, e.g., Ding and VanderWeele 2016 (DOI: 10.1097/EDE.0000000000000457) and references herein."	The use of shifted hypothesis testing is an accepted method to ensure formally that a treatment effect exceeds a certain threshold (e.g., in noninferiority trials). This approach is described as an option. We do agree that methods of sensitivity analysis are available and may also be used in this scenario. We have inserted text to this effect. Thank you for this suggestion.
Liebenhoff, BAH	5	110 - 111	"If indirect comparisons are required, only adjusted indirect comparisons respecting randomisation are appropriate, which means that the evidence network has to be connected." Especially in the case of Orphan Drugs or paediatric medicinal products, it should be ensured, that all available sources of evidence are considered.	Thank you for your comment. We have changed the wording for more clarity.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			So, we suggest replacing the sentence with: "Adjusted indirect comparisons respecting randomisation are more appropriate. Resulting uncertainties of e.g. unconnected networks should be discussed."	
GSK	5	111-112	There is a statement to reflect adjusted indirect comparisons can only be achieved through networks that are connected. Yet later in this section there is a recognition where networks may be disconnected and the methods that have been developed to derive estimates for indirect comparisons are described. Could the sentence on Lines 111-112 reflect a preference for connected nettworks rather than a requirement as the latter seems inconsistent with the guidance acknowledging other methods have been developed for different scenarios including where the network is not connected.	Thank you for your comment. We have changed the wording for more clarity.
GSK	5	121-122	Are there further expectations from the chosen criteria? Apart from being pre-specified transparently?	The summary of the document is not the place to describe these details.
Silke Walleser Autiero Medtronic	5	119-120	Ideally these comparisons are thoroughly investigated in conjunction with subject matter experts e.g. by asking them what they consider to be the most important treatment effect modifiers and making sure they understand what a TEM actually is.	We agree that subject matter experts (often clinical experts) are helpful in determining treatment effect modifiers.
James Ryan AZ	5	128-129	The draft guideline states that: "In the case of disconnected networks (e.g., single-arm trials) and any situations with non randomised trials complete access to the individual patient-level data is required in order to apply methods that can adequately adjust for confounding." While it would be preferrable to have IPD for both comparators, most of the times manufacturers will only have IDP from their own trials. In this case a mix of IPD and aggregate data would be considered, using population-adjusted methods, as MAIC and STC.	The limitations associated with the suggested methods are outlined in the guideline
EFSPI	5	110-111	Current wording: "If indirect comparisons are required, only adjusted indirect comparisons respecting randomisation are appropriate […]"	See general response

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			The explicit recommendation against unanchored indirect comparisons conflicts with the exposition on page 26, lines 762-765. In general, the draft guidance should promote an inclusive approach to using the best available evidence. Recommended rewording: "If indirect comparisons are required, adjusted indirect comparisons respecting randomisation are preferred which means that the evidence network has to be connected. Comparisons in disconnected networks require special considerations and are at higher risk of bias and their use should be justified."	
EFSPI	5	121-122	Current wording: "The model and covariate selection strategies for adjustment must be pre-specified and based on transparent criteria" Pre-specification of effect modifiers helps with addressing multiplicity issues (when accompanied by a clinical rationale) but it is not clear how to operationalize full pre-specification in the context of the JCA. Effect modifiers of interest depend on the pair of treatments being compared and the scale used to measure the releative effectiveness, which again depends the scope of the JCA. The scope will in practice not be fully known at the time phase 3 clinical trials are designed or before the start of the additional analysis. In addition, what is the implication if some effect modifiers are not pre-specified but are nonetheless important to support a JCA? Recommended rewording: "The strategy for eliciting covariates to adjust for must be justified based on clinical and statistical considerations, and accompanied by argumentation on how/why lack of adjustment for these covariates can be expected to lead to bias"	JCA will be an assessment of evidence provided by HTDs. Pre-specification refer to as pre-specification according to the study protocol and statistical analysis plan of a given study. Regarding alignment between development and HTA needs, JSCs are a place to discuss these aspects.
James Ryan AZ	5	111	"If indirect comparisons are required, only **adjusted** indirect comparisons respecting randomisation are appropriate". Issue: the word adjusted can mean different things depending on the technique being considered. Although the original Bucher technique talks about it being adjusted, it is not adjusting for	While we understand adjusted is a term with some ambiguity, it is proposed here as opposed to naïve indirect comparison which is an opposition that is commonly understood.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			population differences etc… that the word is commonly used for now. Suggestion: remove the word adjusted as not all techniques adjust for factors. This will avoid confusion and the remaining statement maintains the same key message.	
James Ryan AZ	5	117	"these methods require that all effect modifiers relevant for adjustment are measured". This presumes that we have an anchored indirect comparison where we perform population adjustment. For an unanchored analysis we should also adjust for prognostic variables. Suggestion: "these methods require that all effect modifiers, **and where appropriate prognostic variables,** relevant for adjustment are measured."	This paragraph concerns connected evidence networks and therefore knowledge of prognostic variables is not generally required.
M. Ermisch – GKV-Spitzenverband	5	124	The concept of "shifting hypothesis" should be explained and its relation e.g. to "important difference"-concepts should be clarified. Criteria and procedural guidance should be given regarding its proposed application. Given the risk of unobserved confounding that is inherent to indirect comparisons, penalising all studies performing indirect comparisons with a shifted null-hypothesis should be further investigated.	An explanation is provided in Section 5.3.
GSK	5	124	Regarding testing shifting hypotheses, where there are great uncertainties with population-adjusted methods, this will in itself lead to an increased hurdle in demonstrating a treatment effect excluding a zero (no different) effect as more variability will be introduced when deriving estimates of treatment effects. Requiring confidence intervals for a treatment effect to be completely above or below a certain threshold may not be possible nor achievable given the clinical trials that form the basis for the indirect comparison were never designed with this in mind. Could there be a recognition that in some situations there may be a need for a larger treatment effect	The use of shifted hypothesis testing is described as an option. It is misleading to describe this approach as "increased hurdle". In fact, it represents an additional option to demonstrate effectiveness of treatments in difficult data situations, in which otherwise no reliable methods are available.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			to be demonstrated more than a zero effect and whether an application would need to meet a higher threshold would be discussed as part of the scoping process.	
GSK	5	124	Suggestion for rewording (additions shown in capitals): "This means that a conclusion ABOUT THE STATISICAL SIGNIFICANCE can be drawn..."	The purpose of a shifted null hypothesis is to test the robustness of a result against residual confusion, and less about statistical significance in itself.
Silke Walleser Autiero Medtronic	5	124	We would welcome a clear definition of the term 'shifted hypothesis', together with an example. From context, I'm guessing that this is simply a non-zero (or for odds ratios & hazard ratios, non-unity) null, e.g., H0: TE<=2 vs H1: TE>2 for an OR...? Referring to the term 'large' ideally should be reconsidered, as it is subjective/relative. It is better to expect an analysis to sufficiently justify if the necessary assumptions for a valid (unbiased) estimate are satisfied. What those entail will vary depending on the analysis but just finding a 'large' result is insufficient.	A definition of a shifted null hypothesis is given in section 5.3. There is no consensus on a specific value for considering an effect sufficiently large for discarding the possibility of residual confounding, hence the qualitative description only.
GSK	5	127	The suggestion for disconnected networks and unanchored comparisons is not always feasible (i.e. ipd not readily available). Is there value in doing MAIC in the absence of ipd?	The limitations associated with MAIC in the absence of IPD is outlined in the guideline.
Matias Olsen, EUCOPE	5	128	In the case of disconnected networks (e.g. single-arm trials), it may not be feasible to have access to complete individual patient-level data, if the comparator study is not owned by the submitting manufacturer. It would be critical to have guidance on what to do in this circumstance and if unanchored population adjustment may be considered.	The document does not describe methods on the basis of available data; rather it describes the limitations that may arise when applying certain methods to certain evidence types. We do not consider the availability of data to be relevant in the determination of methodological rigour.
Prof. Matthias P. Schönermark, M.D., Ph.D. and Svenja Sake, Ph.D. (SKC)	6	Scope and terminology	Comment: A table providing an overview of definitions needed for this document would be very helpful. We recommend to include such a table.	A list of definitions will be provided in the upcoming Practical Guideline: Direct and Indirect Comparisons (deliverable D4.3.1).
Bayer	6	157-	Lines 157-159 state that "evidence must be relevant for the research	These features of population-adjusted

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
		159/I Introduction, objective and scope	question and in most cases should be formulated according to the PICO (Population, Intervention, Comparator, Outcome) framework". The importance of PICO is highlighted again in lines 219-221. Undeniably, choosing the target population of the analysis (the "P" in PICO) is an important aspect when establishing the research question. Nevertheless, conventional evidence synthesis methods (Sections 5.1 and 5.2) produce estimates that are not directly interpretable in any specific target population for decision-making (Manski 2019, Dahabreh et al. 2020, Barker et al. 2022). For example, standard fixed-effects NMA pools study-specific estimates by weighting these according to their precision, without bearing in mind their relevance to a target population of substantive interest. • *Manski, C.F., 2019. Meta-analysis for medical decisions. National Bureau of Economic Research.* • *Dahabreh, I.J., Petito, L.C., Robertson, S.E., Hernán, M.A. and Steingrimsson, J.A., 2020. Toward causally interpretable meta-analysis: Transporting inferences from multiple randomized trials to a new target population. Epidemiology, 31(3), pp.334-344* • *Barker, D.H., Dahabreh, I.J., Steingrimsson, J.A., Houck, C., Donenberg, G., DiClemente, R. and Brown, L.K., 2022. Causally interpretable meta-analysis: Application in adolescent HIV prevention. Prevention Science, 23(3), pp.403-414.* On the other hand, population-adjusted indirect comparisons and meta-regression approaches explicitly specify a target population. As a result, the development of these approaches is important. Lines 718-720 state that these methods "are useful in situations in which an NMA is performed but there is some doubt regarding whether the similarity assumption is valid for some effect modifiers". Arguably, these methods are useful more generally, due to their ability to synthesize information in specific populations. This is particularly the case for meta-regression approaches such as ML-NMR, which have the advantage of estimating treatment effects in any target	methods are described in the sections 5.3.1 to 5.3.3; more details are given in the Practical Guideline (deliverable 4.3.1).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			population.	
GSK	6	151-153	Could a hierarchy of value of evidence be clearly presented in a schematic or Table? Are all analyses of some value in the absence of higher level of evidence? Is there a cut off beyond which the results will not be considered (for example unanchored using aggregate data for adjustments).	It is not possible to develop a simple Table with a hierarchy of value of evidence, because the value of evidence is dependent on too many factors.
Silke Walleser Autiero Medtronic	6	172-174	This language implies that only certain methods can be used for specific situations. There may be scenarios where other methods (potentially new methods) can be applied and there may be times when methods outlined here cannot. Consider changing this to: "which methods are most likely applicable in a particular situation".	We do not think this suggestion will impact the content of the document.
EFSPI	6	165-67	Current wording: "The objective of this document is to describe the methods currently available for direct and indirect treatment comparisons regarding their underlying assumptions, strengths and weaknesses." Since the methodology in this area is constantly evolving, it is key to ensure that the draft guidance remains topical also in the years to come. To ensure this, we recommend to refrain from rejecting methodologies but instead promote an inclusive approach, particularly towards emerging methodologies. Application of emerging methodologies should be accompanied by careful, targeted sensitivity analyses. In addition, we recommend regular updating of the guideline in collaboration with key stakeholders, including academia, and industry experts.	Review period and process will be identified under the CG
M. Ermisch – GKV-Spitzenverband	6	154	The statement: "A systematic literature search is a prerequisite before conducting an evidence synthesis." might lack generality in this context, in particular vis-à-vis the subsequent reference to Art. 9 of the EU regulation. The appropriate general statement here might refer to the prerequisite that all efforts have been made to secure the inclusion of all relevant trials/studies of which the "systematic literature research" should be part. The potential problem of publication bias is addressed in section 3.2	We reworded the statement.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			(p. 15) of the draft document. Here (p. 6) or on p. 15 it should be made clear that in the process of joint assessments the utmost effort to avoid a biased evidence base in terms of trial (results) not having been available should be stressed.	
Matias Olsen, EUCOPE	6	154	Is there a requirement of timing for SLR to be conducted, e.g. maximum 6 months before submission?.	SLR guidance is outside the scope of this guideline.
Matias Olsen, EUCOPE	6	162	"…Consistency in outcome assessment between studies must be checked and discussed". Please specify when in the procedure this will be done and by whom.	We have clarified this paragraph describes the process before the conduct of any statistical analysis of a given evidence synthesis study.
Sebastian Werner vfa	7	187 / I	Please add "with at least 2 studies" at the end of the sentence	It is implied a synthesis of studies cannot be performed without at least two studies to be synthesized.
Mihai Rotaru - EFPIA	7	201 / I	**Direct versus indirect comparisons** Current wording: "However, results from indirect comparisons generally have greater uncertainty than results from direct comparisons. Therefore, direct comparisons based on adequate RCTs with low Risk of Bias (RoB) should be applied whenever possible." Suggested rewording: "However, results from indirect comparisons generally have greater uncertainty than results from direct comparisons. ~~Therefore, direct comparisons based on adequate RCTs with low RoB should be applied whenever possible~~." [note: ~~strikethrough~~ denotes proposed deletion] Rationale: EFPIA requests the deletion of this statement for the following reasons. Firstly, the proposed scoping process (D4.2) of consolidating a PICO based on a survey from EU27 MS will likely result in multiple comparators, and potentially subgroups requested	Thank you for your observation. We don't believe this constitutes a value judgement; rather it reflects the scientific assessment that every assessor should strive towards i.e. to use the most robust evidence with the lowest risk of bias.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			for the JCA. With this approach, conducting indirect treatment comparisons will become a critical and integral part of a JCA. Secondly, EFPIA believes that the statement places indirect treatment comparisons into question, which may be deemed to implicitly incorporate a value judgement as opposed to a representing a scientific assessment based on the available evidence. The acceptable level of uncertainty is a value judgement and should be determined by each MS individually, as stated in the EU HTA regulation, '*It is necessary therefore that Union action is limited to those aspects of HTA that relate to the joint clinical assessment of a health technology and to ensure in particular that there are no value judgements in joint clinical assessments in order to respect the responsibilities of Member States pursuant to Article 168(7) TFEU (L 458/3, 14).*'[1] References: 1. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	7	201 / I	Current wording: "However, results from indirect comparisons generally have greater uncertainty than results from direct comparisons. Therefore, direct comparisons based on adequate RCTs with low Risk of Bias (RoB) should be applied whenever possible." Suggested rewording: "However, results from indirect comparisons generally have greater uncertainty than results from direct comparisons. ~~Therefore, direct comparisons based on adequate RCTs with low RoB should be applied whenever possible~~." [note: ~~strikethrough~~ denotes proposed deletion]	Thank you for your observation. We don't believe this constitutes a value judgement; rather it reflects the scientific assessment that every assessor should strive towards i.e. to use the most robust evidence with the lowest risk of bias.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Rationale: Takeda requests the deletion of this statement for the following reasons. Firstly, the proposed scoping process (D4.2) of consolidating a PICO based on a survey from EU27 MS will likely result in multiple comparators, and potentially subgroups requested for the JCA. With this approach, conducting indirect treatment comparisons will become a critical and integral part of a JCA. Secondly, we believe that the statement places indirect treatment comparisons into question, which may be deemed to implicitly incorporate a value judgement as opposed to a representing a scientific assessment based on the available evidence. The acceptable level of uncertainty is a value judgement and should be determined by each MS individually, as stated in the EU HTA regulation, '*It is necessary therefore that Union action is limited to those aspects of HTA that relate to the joint clinical assessment of a health technology and to ensure in particular that there are no value judgements in joint clinical assessments in order to respect the responsibilities of Member States pursuant to Article 168(7) TFEU (L 458/3, 14).*'[1] References: 10. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
MTE	7	204-207	For simplicity, we use effectiveness as the common term to describe efficacy or effectiveness throughout the rest of this document. Effectiveness also includes safety within the context of this document. Furthermore, treatment, intervention and health technology are all terms used for any health technology that can be assessed.	Thank you but we consider that this comment is outside the scope of this guideline.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			For medical technologies, the approach to consider the relative safety of a complex intervention as part of the relative effectiveness will be an excellent way forward to appreciate the relative effects. This to avoid also the ambiguity of a device related safety (covered by the MDR/IVDR) and the complex intervention which for the targeted technologies is implanted and have an effect related not only to the technologies but as well by the users. A relative effect on the save use and patient safety will provide relevant information of this domain of the effectiveness. As for the efficacy and effectiveness, we do call for to further describe opportunties to determing relative efficacy and relative effectiveness wherebhy also the stenght of trial design might be better suited for one or the other. As for the RCT, we would call to further elaborate on the type of randomization whereby further attention to be given for Pragmatic Randomization with open label blinded endpoints design which increases the ability of randomization in implantable devices or treatment whereby double blinding is challenging. Within this context we would also welcome a further elaboration on novel trial design and advanced analysis techniques. Beyond the relative effects it might as well be important to put forward some consideration on the difference in absolute effects and the supportive evidence as an alternative way to appreciate the value of a technology.	
Silke Walleser Autiero Medtronic	7	201-202	The term 'certainty' should be changed to 'validity'. This occurs throughout the document also.	Validity is only a part of certainty of results. Certainty as a more general term is a better fit.
GSK	7	187	Please add "with at least 2 studies" at the end of the sentence	We think this addition is not required.
M. Ermisch – GKV-Spitzenverband	7	199	We propose to change the sentence by inserting an additional aspect: "For cases in which no data for the relevant direct comparison are available or the research question requires simultaneous comparison	Thank you. We have revised the text accordingly.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			of more than two interventions [suggested insert: which have not been compared directly in trials], methods for indirect comparisons are available." The original sentence does not address the possibility that simultaneous comparisons of more than two interventions are available from (multi-arm) trials and, thus, might not always require the use of indirect comparisons – hence the suggested insert. Although such trials might be very rare up to today, this could – and in some cases certainly should – change in the future, e.g. in the form of "platform trials", which have achieved a certain prominence in the pandemic. Even if comparisons in platform trials may be indirect comparisons, they can allow for a pronounced reduction of heterogeneity.	
Sarah Smith, Lumanity	7	Line 203	The definition of "adequate" is not clear here. Please clarify what is meant by "adequate" or if referring to low RoB then suggest removing the word adequate.	Adequate is defined within the summary: appropriately designed and with low RoB.
vfa Sebastian Werner vfa	8	235 – 248 / II.1	Please add the concept of predictive variable.	We think the term effect modifier covers the concept of predictive variable in our document.
GSK	8	235 - 248	Please add the concept of predictive variable.	We think the term effect modifier covers the concept of predictive variable in our document.
Advanced Medical Services GmbH	8	219-222	Reference is made to two guidelines (D4.3.1 and D4.6.1), which are currently not accessible. At least the rough ideas of these guidelines should be quoted.	Thank you. These will be available within the timeframes outlined on the EUnetHTA website and therefore it is not proposed to add additional information here.
Silke Walleser Autiero Medtronic	8	232-234	May not be achieved. Just as in an RCT there could be bias in the unobserved and just by chance there could be balance in the unobserved using other methods, it is difficult to ascertain. The key here is that it can never be assessed. (If it wasn't achieved PSM it will not be valid).	We agree with the remark, but it does not contradict the general statement made in the document.
EFSPI	8	240-242	Current wording: "Thus, effect modifiers are specific to the pair of treatments being compared and to the scale used to measure the	Thank you. We have amended the document for clarity.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			relative treatment effectiveness. An example is the stage of a particular disease [...]" It is not correct that disease stage is generally an effect modifier. To avoid future assessors insisting that disease stage must be considered an effect modifier, we recommend removing (or at least clarifying that disease stage is not always an effect modifier). Recommended rewording: "For example, in some particular setting, the stage of a particular disease is an effect modifier: the relative effectiveness of the treatment being studied to its comparator is not the same for patients at an early stage and patients at a later stage of the disease."	
Roche	8	243-244/1	The proposed guidance states *"effect modifiers can be considered interaction terms between the treatment and the outcome of interest".* We propose replacing this definition with the definition given in the reference [17, page 227]: *"[...] effect-modifying covariates, in other words of interactions between the treatment effect and trial-level or patient-level variables".* [17] Dias S, Ades A, Welton N, et al. Network Meta-Analysis for Decision Making. Chichester, UK: Wiley; 2018.	Thank you for your comment. There has been much discussion around the terminology and definitions, and we believe the definition currently in the text is appropriate in the context of the document.
Sarah Smith, Lumanity	8	Line 245/246	Referring to the wording "…although in general not all prognostic variables will be effect modifiers", it is also the case that treatment effect modifiers are not necessarily prognostic variables. Therefore, we suggest changing the sentence to "…although in general not all prognostic variables will be effect modifiers and *vice versa*."	It is possible in theory that a variable is an effect modifier but not a prognostic variable (in the case of a cross-over interaction), but this situation is impossible in adequately conducted RCTs and therefore, this plays no role here.
Mihai Rotaru - EFPIA	8	243 / II.1	**Editorial: Effect-modifying covariates** <u>Current wording:</u> "Effect modifiers can be considered interaction terms between the	We changed the wording of this sentence.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			treatment and the outcome of interest." Suggested rewording: "Effect**-modifying** ~~modifiers~~ **covariates** can be considered **as** interaction**s** ~~terms~~ between the treatment and the outcome of interest **effect and trial-level or patient-level variables [17].**" [note: **bold** and ~~strikethrough~~ denotes proposed inclusion and deletion, respectively] Rationale: EFPIA believes the definition of effect modifiers in the current guidance seems misleading. We recommend using the definition from the textbook (Dias, 2018[1]) to improve clarity. References 1. Dias S, Ades A, Welton N, et al. Network Meta-Analysis for Decision Making. Chichester, UK: Wiley; 2018. Available at: https://www.wiley.com/en-us/Network+Meta+Analysis+for+Decision+Making-p-9781118647509.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	8	243 / II.1	Current wording: "Effect modifiers can be considered interaction terms between the treatment and the outcome of interest." Suggested rewording: "Effect**-modifying** ~~modifiers~~ **covariates** can be considered **as** interaction**s** ~~terms~~ between the treatment and the outcome of interest **effect and trial-level or patient-level variables [17].**" [note: **bold** and ~~strikethrough~~ denotes proposed inclusion and deletion, respectively] Rationale:	We changed the wording of this sentence.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			The definition of effect modifiers in the current guidance is misleading. We recommend using the definition from the textbook (Dias, 2018[1]) to improve clarity. References 2. Dias S, Ades A, Welton N, et al. Network Meta-Analysis for Decision Making. Chichester, UK: Wiley; 2018. Available at: https://www.wiley.com/en-us/Network+Meta+Analysis+for+Decision+Making-p-9781118647509.	
Mihai Rotaru - EFPIA	8	246 / II.1	**Editorial: Population adjusted methods – Confounders** Current wording: "In the context of a comparison between two treatments, a confounder is a characteristic that affects both the treatment received and the outcome; in other words, a prognostic variable that is not "balanced" between treatment groups." Suggested rewording: "In the context of a comparison between two treatments, a confounder is a characteristic that affects both the likelihood of receiving the treatment **and the outcome of the treatment given**~~; in other words, a prognostic variable that is not "balanced" between treatment groups~~." [note: **bold** and ~~strikethrough~~ denotes proposed inclusion and deletion, respectively] Rationale: EFPIA believes that a confounder is not just an imbalanced prognostic factor. If treatment decisions are not affected by this factor, the factor is not a confounder. In addition, confounders can be also predictive factors without prognostic value, for an example, biomarkers. Therefore, we propose that the sentence should be	Thank you. We have revised the sentence for clarity.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			modified as suggested above.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	8	246 / II.1	Current wording: "In the context of a comparison between two treatments, a confounder is a characteristic that affects both the treatment received and the outcome; in other words, a prognostic variable that is not "balanced" between treatment groups." Suggested rewording: "In the context of a comparison between two treatments, a confounder is a characteristic that affects both the likelihood of receiving the treatment **and the outcome of the treatment given**~~; in other words, a prognostic variable that is not "balanced" between treatment groups~~." [note: **bold** and ~~strikethrough~~ denotes proposed inclusion and deletion, respectively] Rationale: A confounder is not just an imbalanced prognostic factor. If treatment decisions are not affected by this factor, the factor is not a confounder. In addition, confounders can be also predictive factors without prognostic value, for an example, biomarkers. Therefore, we propose that the sentence should be modified as suggested above.	Thank you. We have revised the sentence for clarity.
Bayer	9	254-258/ Section 1	This text discusses the summary effect measure in meta-analysis, stating that the summary statistic computed for each study should be the same. An important point that should be raised by the document is the non-collapsibility of some of these effect measures, e.g., the hazard ratio and the odds ratio. For non-collapsible effect measures, marginal and conditional estimands do not coincide, even in the absence of confounding. In addition, conditional estimands conditioning on different covariate sets do not coincide either. Plainly speaking, marginal odds ratios have to be combined with marginal	Thank you for this comment; however, these details would be misplaced in this short section. The issue of marginal vs. conditional estimands will be discussed in the Practical Guideline (deliverable 4.3.1).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			odds ratios; conditional hazard ratios have to be combined with conditional hazard ratios, adjusted for the same covariate sets. Comparisons must involve compatible estimates (Daniel, Zhang and Farewell, 2021), otherwise there is a risk of bias (Remiro-Azócar, Heath and Baio, 2021). For reimbursement decisions at the population level, marginal estimates should be combined as the target should be a marginal estimand (Remiro-Azócar, 2022 and related discussion). • *Daniel, R., Zhang, J. and Farewell, D., 2021. Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. Biometrical Journal, 63(3), pp.528-557.* • *Remiro‐Azócar, A., Heath, A. and Baio, G., 2021. Methods for population adjustment with limited access to individual patient data: A review and simulation study. Research synthesis methods, 12(6), pp.750-775.* *Remiro-Azócar, A., 2022. Target estimands for population-adjusted indirect comparisons. In press, Statistics in Medicine https://arxiv.org/pdf/2112.08023.pdf*	
James Ryan AZ	9	258-262	Add in text in bold: Indirect evidence cannot ensure balance of both known and unknown effect modifiers to the same degree as direct evidence from RCTs and, all else being equal, is more uncertain as a result. However, when direct evidence informing a comparison of interest is not available, comparisons using indirect evidence need to be made, **even when heterogeneity exists between study population studied or only where single arm studies are available.**	We disagree with the addition. If hetergeneity is too large, no evidence synthesis (neither a meta-analysis nor a NMA) should be performed (see Key Points I).
Sebastian Werner vfa	9	253-255 / II.1	"This very high RoB is likely to carry through and can be compounded when combining evidence from these sources." Please consider adding the following: "In special situations (e.g., challenges of patient recruitment for orphan diseases or in small and vulnerable populations) single-arm trials, cohort studies, case-control studies, other observational	Thank you we don't believe the sentence requires that specific situations such as described are needed to improve clarity of the statement.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			studies, and the use of historical controls might reflect the best available evidence due to practical and ethical considerations. Appropriate approaches must be used and results have to be discussed carefully due to uncertainties. Additionally, the high RoB can be compounded when combining evidence from these sources."	
Edwards Lifesciences	9	261-262/ Section 1 Types of evidence	We propose to change the sentence as follows (additional text in bold): "However, when direct evidence informing a comparison of interest is not available **or is not ethical**, comparisons using indirect evidence need to be made".	The discussion of what is and what is not ethical in terms of randomisation can be highly debated and remains controversial.
Mihai Rotaru - EFPIA	9	253 / II.1	**Context: subjective text/ value judgement in risk of bias language** <u>Current wording:</u> "This very high risk of bias (RoB) is likely to carry through and can be compounded when combining evidence from these sources." <u>Suggested rewording:</u> ""This ~~very high~~ risk of bias (RoB) is likely to carry through and can be compounded when combining evidence from these sources." [note: ~~strikethrough~~ denotes proposed deletion] <u>Rationale:</u> EFPIA acknowledges that nonrandomised evidence can have a risk of bias versus a RCT. However, the term 'very high' is a subjective assessment (and as stated in an earlier comment) can bias assessors of JCAs from the onset against nonrandomised evidence. Furthermore, there are established methods and guidelines that have been developed to reduce the risk of bias when non-randomised evidence is used for comparative effectiveness assessments. As such, EFPIA recommends: 2. Deletion of the term 'very high' from the statement,	Thank you. We have softened the wording.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			3. Acknowledge that although there is a risk of bias when non-randomised evidence is used, there are methods and guidelines that provide recommendations and methods to reduce the risk of bias.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	9	253 / II.1	Current wording: "This very high risk of bias (RoB) is likely to carry through and can be compounded when combining evidence from these sources." Suggested rewording: ""This ~~very high~~ risk of bias (RoB) is likely to carry through and can be compounded when combining evidence from these sources." [note: ~~strikethrough~~ denotes proposed deletion] Rationale: We acknowledge that nonrandomised evidence can have a risk of bias versus a RCT. However, the term 'very high' is a subjective assessment and can bias assessors of JCAs from the onset against nonrandomised evidence. Furthermore, there are established methods and guidelines that have been developed to reduce the risk of bias when non-randomised evidence is used for comparative effectiveness assessments. The following is a list of existing HTA methods guides which all state that ITCs can be conducted by different ITC methods and propose methods to reduce the risk of bias: **Sweden** Dental and Pharmaceutical Benefits Agency (TLV) (2017). www.tlv.se/lakemedel/ansok-om-pris-eller-subvention/ **The Netherlands** Zorginstituut Nederland (2017).	Thank you. We have softened the wording.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			www.zorginstituutnederland.nl/over-ons/publicaties/rapport/2016/09/09/procedure-beoordeling-extramurale-geneesmiddelen Zorginstituut Nederland (2017). www.zorginstituutnederland.nl/publicaties/rapport/2015/01/15/beoordeling-stand-van-de-wetenschap-en-praktijk **France** Haute Autorite de Sante. Depot de dossier de transparence (2017). www.has-sante.fr/portail/jcms/c 1046750/f r/depot-de-dossier-de-transparence **Canada** CADTH Common Drug Review Committee (2017). www.cadth.ca/about-cadth/what-we-do/products-services/cdr/common-drug-re view-submissions/guidelines-procedures-templates **United Kingdom** NICE Decision Support Unit (DSU) guidance: -https://nicedsu.sites.sheffield.ac.uk/tsds/evidence-synthesis-tsd-series -https://nicedsu.sites.sheffield.ac.uk/tsds/multivariate-meta-analysis-tsd -https://nicedsu.sites.sheffield.ac.uk/tsds/observational-data-tsd -https://nicedsu.sites.sheffield.ac.uk/tsds/population-adjusted-indirect-comparisons-maic-and-stc). **Australia** Pharmaceutical Benefits Advisory Committee (2017). https://pbac.pbs.gov.au/ In line with the HTA methods guidance from many reputable HTA	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			agencies, Takeda recommends the following changes be made to the existing guideline: 4. Deletion of the term 'very high' from the statement, 2. Acknowledge that although there is a risk of bias when non-randomised evidence is used, there are methods and guidelines that provide recommendations and methods to reduce the risk of bias.	
M. Ermisch – GKV-Spitzenverband	9	261	The necessity of using indirect comparisons should not be stated in absolute terms. A wording like "need to be tried" instead of "is necessary" seems more appropriate. The qualification of "necessity" seems to arise as, come what may, a relative effect estimate is required and no direct comparisons available. However, what is, e.g., to be done, if not only direct A-B-estimates are lacking but also no common "C" is to be found? Additionally, it should be noted that in other scenarios (insurmountable heterogeneity, section 3.3) it seems clear that situations are possible in which no effectiveness estimate may be derived.	Thank you. We have amended the text accordingly.
Advanced Medical Services GmbH	10	264-270	During the first two years of HTAR (2025-2027), the focus will be on cancer drugs and ATMPs. It should be discussed in how far Networks of Evidence can be generated for orphan drugs and/or ATMPs.	This is out of the scope of this Guideline.
Sebastian Werner vfa	10	Figure 1 / II.2	From an evidence level perspective, it would make more sense to start with network b). In network c) also indirect comparisons are possible. Therefore, dotted lines should be added.	The first figure introduces the simplest case for indirect evidence as described above. For Figure C, adding a lot of dotted line would alter the readability of the document, but we have specified within the legend that indirect comparisons can be made.
Sarah Smith, Lumanity	10	Line 286/287	Suggest referring to Figure 1 to help define "more complex NMA methods": "In more complex connected networks of evidence (as illustrated in Figure 1b and 1c), more complex NMA methods (Section 5.2) are needed."	Thank you. We have revised the text accordingly.
GSK	10	Figure 1	From an evidence level perspective, it would make more sense to	The first figure introduces the simplest

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			start with network b). In network c) also indirect comparisons are possible. Therefore, dotted lines should be added.	case for indirect evidence as described above. For Figure C, adding a lot of dotted line would alter the readability of the document, but we have specified within the legend that indirect comparisons can be made.
M. Ermisch – GKV-Spitzenverband	10	279 ff.	It is difficult to understand where the authors actually distinguish between high- and low-quality evidence. In our view, it should be clearly stated that the line is between randomised and non-randomised comparisons – which would imply that it is no option to connect an otherwise unconnected network via a parallel cohort study with 2 (or more) interventions without randomisation. Hence, NMA should only be done in networks of RCTs and should only be used if transitivity is not in doubt.	We think the section "network of evidence" clearly states that connected networks implies connected networks by means of RCTs. Regarding the issues of connecting an otherwise unconnected network via a cohort study, we do think the last paragraph of the section tackles the issue appropriately.
Silke Walleser Autiero Medtronic	11	294-310	Disconnected NMA methodology is advancing. We suggest to consider including updated references to this approach: eg. *Med Decis Making*. 2022 May 7; doi: 10.1177/0272989X221097081.	Thank you for this suggestion - we have added a reference to this paper.
BIOTRONIK SE & Co. KG	11	310-312	'*… because the EU regulation requires comparative results on the basis of adequate comparisons (PICO framework).'* The HTA R does mention direct comparative evidence twice and in direct relation to pharmaceuticals ('medicinal products'). It does not specify such need for non-pharmaceuticals, allowing in principle for other types of evidence to be considered. It would be interesting to see a proposal for appropriately considering and analysing such other evidence. The HTA R makes no reference to the PICO framework, so it is unclear why this is mentioned here. We suggest that this reference is deleted.	This sentence has been revised according to a comment from GKV-Spitzenverband.
EFSPI	11	310-312	Current wording: "The use of such evidence in JCA is highly problematic because the EU regulation requires comparative results on the basis of adequate comparisons" The statement suggests that disconnected networks may not be accepted, conflicting with the exposition lines 314-320 in the guideline and HTAR (L458/6, art 18 para 4) which includes the	The statement is factual in terms of the methodological rigour – the acceptability of such evidence will be taken at member state level.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			provision 'when appropriate'. Generally, there should be an inclusive approach towards the available evidence, and in for example rare diseases, it may be the only available evidence. Recommended rewording: "The use of such evidence in JCA requires careful argumentation and appropriate sensitivity analyses as the EU regulation is preferential towards comparative results being the basis of adequate comparisons"	
Mihai Rotaru - EFPIA	11	289 / 2	**Editorial: Strong equivalence assumption** <u>Current wording:</u> "(b) Two RCTs (A vs B, C vs D) for which the required comparison is C versus B, so a strong assumption must be made in relation to the equivalence of the comparisons and the relative treatment effectiveness". <u>Suggested rewording:</u> "(b) Two RCTs (A vs B, C vs D) for which the required comparison is C versus B, so a**n** ~~strong~~ assumption must be made in relation to the equivalence of the comparisons and the relative treatment effectiveness". [Note: **bold** and ~~strikethrough~~ denotes proposed inclusion and deletion, respectively] <u>Rationale:</u> EFPIA suggests removing the word 'strong' as it is subjective; since there could be instances where the available data does not invalidate the assumption of equivalence.	If there are strong arguments that this strong assumption can be made, the corresponding analysis would be valid. Nevertheless, it remains that the assumption is a strong one, therefore we consider that it is important to reiterate it in conjunction with the figure
Tanja Podkonjak – Takeda Pharmaceuticals International	11	289 / 2	<u>Current wording:</u> "(b) Two RCTs (A vs B, C vs D) for which the required comparison is C versus B, so a strong assumption must be made in relation to the equivalence of the comparisons and the relative treatment effectiveness".	If there are strong arguments that this strong assumption can be made, the corresponding analysis would be valid. Nevertheless, it remains that the assumption is a strong one, therefore

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
AG			Suggested rewording: "(b) Two RCTs (A vs B, C vs D) for which the required comparison is C versus B, so a**n** ~~strong~~ assumption must be made in relation to the equivalence of the comparisons and the relative treatment effectiveness". [Note: **bold** and ~~strikethrough~~ denotes proposed inclusion and deletion, respectively] Rationale: Takeda suggests removing the word 'strong' as it is subjective; since there could be instances where the available data does not invalidate the assumption of equivalence.	we consider that it is important to reiterate it in conjunction with the figure
Mihai Rotaru - EFPIA	11	297 / 2	**Networks of evidence: Disconnected evidence networks (1)** Current wording: "Disconnected networks such as those illustrated in Figure 2 are problematic since there is no way in which the comparators of interest can be compared using paths involving evidence from randomised or comparative trials." Suggested rewording: "In disconnected networks such as those illustrated in Figure 2 ~~are problematic since~~ there is no way in which the comparators of interest can be compared using paths involving evidence from randomised or comparative trials. This implies that assumptions about equivalence of the comparisons have to be proven or made" [Note: ~~strikethrough~~ denotes proposed deletion] Rationale: EFPIA believes that describing these analyses and types of evidence as 'are problematic' is subjective and a value judgement which is outside of the scope of the EU HTA regulation.	We describe them as problematic in terms of the methodological rigour. We consider this to be accurate. The acceptability of such networks will differ between member states.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			The current statement stands to potentially bias the future JCA assessors towards only one type of evidence (RCTs) and lead them to disregard or not consider other evidence sources – both are not appropriate as the totality of evidence should be considered for a thorough JCA. EFPIA recommends that the methodological guideline should focus on providing the assessors with a summary of the types of required assumptions and methods, without classifying them subjectively, potentially biasing the perceptions of assessors.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	11	297 / 2	Current wording: "Disconnected networks such as those illustrated in Figure 2 are problematic since there is no way in which the comparators of interest can be compared using paths involving evidence from randomised or comparative trials." Suggested rewording: "In disconnected networks such as those illustrated in Figure 2 ~~are problematic since~~ there is no way in which the comparators of interest can be compared using paths involving evidence from randomised or comparative trials. This implies that assumptions about equivalence of the comparisons have to be proven or made" [Note: ~~strikethrough~~ denotes proposed deletion] Rationale: Describing these analyses and types of evidence as 'are problematic' is subjective and a value judgement which is outside of the scope of the EU HTA regulation. The current statement stands to potentially bias the future JCA assessors towards only one type of evidence (RCTs) and lead them to disregard or not consider other evidence sources – both are not appropriate as the totality of evidence should be considered for a thorough JCA.	We describe them as problematic in terms of the methodological rigour. We consider this to be accurate. The acceptability of such networks will differ between member states.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Takeda recommends that the methodological guideline should focus on providing the assessors with a summary of the types of required assumptions and methods, without classifying them subjectively, potentially biasing the perceptions of assessors.	
Sebastian Werner vfa	11	301-303 / II.2	"However, these approaches rely on very strong assumptions that need to be examined carefully for any specific application." Please add to the sentence above "In order to use the best available evidence potential limitations can be acknowledged e.g., by conducting various sensitivity analyses to investigate the robustness of the results."	Reference to sensitivity analysis to explore uncertainty is made throughout the document and is also included in the conclusion. We don't believe there is additional need to mention it here.
Roche	11	309-312/2	We agree that disconnected evidence networks come with major challenges and that the field is still evolving. Nevertheless, we should expect to encounter such evidence setups, in particular considering rare cancers, ATMPs, and orphan medicinal products. Therefore, we recommend keeping the first sentence in lines 309-312, but dropping the second one, such that the paragraph ends *"Currently, there is no gold-standard method that addresses the issue of disconnectedness of evidence networks."*	Thank you for your comment – we believe that dropping the second sentence would change the meaning and not be clear therefore we propose to leave the text as is.
Mihai Rotaru - EFPIA	11	309 / 2	**Networks of evidence: Disconnected evidence networks (2)** <u>Current wording:</u> "Currently, there is no gold-standard method that addresses the issue of disconnectedness of evidence networks. The use of such evidence in JCA is highly problematic because the EU regulation requires comparative results on the basis of adequate comparisons (PICO framework)." <u>Suggested rewording:</u> "Currently, there is no gold-standard method that addresses the issue of disconnectedness of evidence networks. **Disconnected networks have limitations compared to RCTs and require certain assumptions which may increase uncertainty in comparative results**. The use of such ~~evidence~~ **networks may be considered** in JCA ~~is highly problematic because the EU regulation~~	These lines have been revised according to a comment from GKV-Spitzenverband.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			~~requires comparative results on the basis of adequate comparisons~~ **however the results should be interpreted with caution and the limitations of the evidence base clearly captured and described."** [Note: **bold** and ~~strikethrough~~ denotes proposed insertion and deletion, respectively] <u>Rationale:</u> EFPIA believes that describing methods used in data availability situations with disconnected networks as 'highly problematic' due to the absence of a gold-standard method is inappropriate. Instances of disconnected networks will be a frequent issue, particularly due to the potential for multiple PICOs (e.g., alternative comparators); whilst also considering the anticipated evidence packages for many ATMPs, orphan and oncology medicines. In these disease areas, it may be often unethical or unfeasible to conduct an RCT or, as in the case of oncology, many variations of standards of care may exist across EU27 MS, with the potential for limited data. EFPIA wishes to highlight this in the context of the EU HTA regulation, which states that, '*Methodologies for performing joint clinical assessments and joint scientific consultations should be adapted to include specificities of new health technologies for which some data may not be readily available. This may be the case for, inter alia, orphan medicinal products, vaccines and advanced therapy medicinal products (L 458/5; Section 24).*' [1] Furthermore, EFPIA believes that describing these analyses and types of evidence as 'highly problematic' is subjective and a value judgement which is outside of the scope of the EU HTA regulation. The current statement stands to potentially bias the future JCA assessors towards only one type of evidence (RCTs) and lead them to disregard or not consider other evidence sources – both are not	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			appropriate as the totality of evidence should be considered for a thorough JCA. EFPIA acknowledges that non-randomised evidence and disconnected networks have more limitations than RCTs and require certain assumptions (which could be demonstrated to hold), however, we suggest that the methodological guideline is amended with regards the statements relating to non-randomised or disconnected evidence while, still listing their required assumptions and limitations. <u>References:</u> 1. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	11	309 / 2	<u>Current wording:</u> "Currently, there is no gold-standard method that addresses the issue of disconnectedness of evidence networks. The use of such evidence in JCA is highly problematic because the EU regulation requires comparative results on the basis of adequate comparisons (PICO framework)." <u>Suggested rewording:</u> "Currently, there is no gold-standard method that addresses the issue of disconnectedness of evidence networks. **Disconnected networks have limitations compared to RCTs and require certain assumptions which may increase uncertainty in comparative results**. The use of such ~~evidence~~ **networks may be considered** in JCA ~~is highly problematic because the EU regulation requires comparative results on the basis of adequate comparisons~~**h and the limitations of the evidence base clearly captured and described."**	These lines have been revised according to a comment from GKV-Spitzenverband.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			[Note: **bold** and ~~strikethrough~~ denotes proposed insertion and deletion, respectively] <u>Rationale:</u> Takeda believes describing methods used in data availability situations with disconnected networks as 'highly problematic' due to the absence of a gold-standard method is inappropriate. Instances of disconnected networks will be a frequent issue, particularly due to the potential for multiple PICOs (e.g., alternative comparators); whilst also considering the anticipated evidence packages for many ATMPs, orphan and oncology medicines. In these disease areas, it may be often unethical or unfeasible to conduct an RCT or, as in the case of oncology, many variations of standards of care may exist across EU27 MS, with the potential for limited data. EU HTA Regulation, states that, *'Methodologies for performing joint clinical assessments and joint scientific consultations should be adapted to include specificities of new health technologies for which some data may not be readily available. This may be the case for, inter alia, orphan medicinal products, vaccines and advanced therapy medicinal products (L 458/5; Section 24).'* [1] Describing these analyses and types of evidence as 'highly problematic' is subjective and a value judgement which is outside of the scope of the EU HTA regulation. The current statement stands to potentially bias the future JCA assessors towards only one type of evidence (RCTs) and lead them to disregard or not consider other evidence sources – both are not appropriate as the totality of evidence should be considered for a thorough JCA. While we acknowledge that non-randomised evidence and disconnected networks have more limitations than RCTs and require certain assumptions (which could be demonstrated to hold), we	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			suggest that the methodological guideline is amended with regards the statements relating to non-randomised or disconnected evidence while, still listing their required assumptions and limitations. References: 2. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Mihai Rotaru - EFPIA	11	314 / 2	**Networks of evidence: Observational studies** Current wording: "Observational studies examining two or more interventions have been used to connect an otherwise disconnected network [65]. This can allow for comparisons that otherwise would not be possible. However, relying solely on these methods to produce an unbiased estimate of the relative effectiveness of a treatment(s) of interest in practical settings remains controversial," Suggested rewording: "Observational studies examining two or more interventions have been used to connect an otherwise disconnected network [65]. This can allow for comparisons that otherwise would not be possible. ~~However, relying solely on these methods to produce an unbiased estimate of the relative effectiveness of a treatment(s) of interest in practical settings remains controversial,~~" [note: ~~strikethrough~~ denotes proposed deletion] Rationale: EFPIA recommends inclusion of additional information to provide clarity for data availability situations where observational studies will be required for HTDs.	We think the statement that these comparisons are controversial is correct.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Tanja Podkonjak – Takeda Pharmaceuticals International AG	11	314 / 2	Current wording: "Observational studies examining two or more interventions have been used to connect an otherwise disconnected network [65]. This can allow for comparisons that otherwise would not be possible. However, relying solely on these methods to produce an unbiased estimate of the relative effectiveness of a treatment(s) of interest in practical settings remains controversial," Suggested rewording: "Observational studies examining two or more interventions have been used to connect an otherwise disconnected network [65]. This can allow for comparisons that otherwise would not be possible. ~~However, relying solely on these methods to produce an unbiased estimate of the relative effectiveness of a treatment(s) of interest in practical settings remains controversial,~~" [note: ~~strikethrough~~ denotes proposed deletion] Rationale: We request the inclusion of additional information to provide clarity for data availability situations where observational studies will be required for HTDs.	We think the statement that these comparisons are controversial is correct.
M. Ermisch – GKV-Spitzenverband	11	310 ff.	The statement that the use of evidence from disconnected trials (e.g. single arm trials, historical controls) in the JCA is problematic "because the EU regulation requires comparative results on the basis of adequate comparisons" is misleading, as it implies that changing the regulation would be a solution to the problem. However, the real problem is that such data do not represent adequate comparisons and, thus, are not able to deliver reliable information on the added benefit of one intervention.	Thank you for your comment. We have revised the text accordingly.
M. Ermisch – GKV-Spitzenverban	11	316	The sentence "This can allow for comparisons…" should be changed to "It has been proposed to use this approach to perform comparisons…"	Thank you. We have revised the text accordingly.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
d			This would better convey the information that the validity of these approaches is questionable, as outlined in the next sentences of the paragraph.	
Bayer	12	320-322/ 2 Networks of evidence	Current: In the context of JCA, this could mean that unreliable evidence from observational studies should not be used because the corresponding results will be highly uncertain and would not provide a meaningful estimate of the relative treatment effectiveness. Proposed: In the context of JCA, this could mean that less reliable evidence from observational studies should be used with caution because the corresponding results could be highly uncertain. Rationale: There might be cases, especially for indirect comparisons, where the only available evidence is based on observational studies. Even in this cases, available evidence should not be completely ignored.	We have altered the wording to read "unlikely to provide a meaningful estimate" rather than "would not provide a meaningful estimate".
Paolo Morgese - ARM	12	320-325	The statement is vague and implies a systematic dismissal of data from observational studies. Given that RCTs are not feasible for many ATMPs, observational studies are an important source of evidence. These guidelines should also provide guidance how to generate reliable data from observational studies.	We have altered the wording to read "unlikely to provide a meaningful estimate" rather than "would not provide a meaningful estimate".
EFSPI	12	320-322	Current wording: "[...] this could mean that unreliable evidence from observational studies should not be used" The draft guidance should promote an inclusive approach to using the best available evidence. If only aggregate observational evidence is available, it should still be acceptablefor the merits of this data to be assessed and where justified for this data to be included. Recommended rewording: "In the context of the JCA, this could mean that evidence from observational studies should be used cautiously and always be accompanied by appropriate sensitivity analyses to explore the impact of data limitations and modelling assumptions on the conclusions".	We have altered the wording to read "unlikely to provide a meaningful estimate" rather than "would not provide a meaningful estimate".
Mihai Rotaru - EFPIA	12	320 / 2	Current wording: "In the context of JCA, this could mean that unreliable evidence from observational studies should not be used because the corresponding	We have altered the wording to read "unlikely to provide a meaningful estimate" rather than "would not

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			results will be highly uncertain and would not provide a meaningful estimate of the relative treatment effectiveness." Proposed wording: "~~In the context of JCA, this could mean that unreliable evidence from observational studies should not be used because the corresponding results will be highly uncertain and would not provide a meaningful estimate of the relative treatment effectiveness.~~ **In situations with limited data or where randomised data are not available, observational studies may be used, however this increases the uncertainty of the results. The uncertainty should be captured and thoroughly explored with sensitivity analyses. The guideline recommends MS apply caution and consider the uncertainty in interpreting the results**.' [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] Rationale: EFPIA recommends that the methodological guideline should define what is considered 'unreliable' in this context. For example, would this apply to all observational, nonrandomised data including real-world evidence (RWE). EFPIA wishes to highlight that RWE has become important for demonstrating effectiveness in the real-world setting, particularly to assess effectiveness, and inform historical controls and address uncertainty. However, the guidance text states that observational data could be used only if IPD are available to allow for rigorous adjustment for confounding. Access to IPD from observational studies and RWE (i.e., registries) may not always be feasible, nor ethical due to data privacy considerations, particularly in rare disease settings. Even if a network meta-analysis or other type of indirect comparison is done using data from RCTs, only variables reported in the studies can be used for adjustment for confounding,	provide a meaningful estimate".

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			or sometimes the only available evidence for a comparator is from a single-arm trial. It would be useful for the methodological guideline to discuss the role and incorporation of real-world evidence (RWE) in JCA. As such, EFPIA is concerned that there may be cases where the only available evidence is based on observational studies, particularly when recognising those included in Phase I and II JCA process (orphan diseases, oncology and ATMPs), given the greater potential for EU marketing authorisation on the basis of assessment with non-RCT evidence (i.e., single arm trials). As such, the primary objective should remain the consideration of the available and not exclusively theoretically best-possible evidence.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	12	320 / 2	Current wording: "In the context of JCA, this could mean that unreliable evidence from observational studies should not be used because the corresponding results will be highly uncertain and would not provide a meaningful estimate of the relative treatment effectiveness." Proposed wording: "~~In the context of JCA, this could mean that unreliable evidence from observational studies should not be used because the corresponding results will be highly uncertain and would not provide a meaningful estimate of the relative treatment effectiveness.~~ **In situations with limited data or where randomised data are not available, observational studies may be used, however this increases the uncertainty of the results. The uncertainty should be captured and thoroughly explored with sensitivity analyses. The guideline recommends MS consider the uncertainty in interpreting the results**.' [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively]	We have altered the wording to read "unlikely to provide a meaningful estimate" rather than "would not provide a meaningful estimate".

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Rationale: We request the methodological guideline define what is considered 'unreliable' in this context. For example, would this apply to all observational, nonrandomised data including real-world evidence (RWE)? RWE has become important for demonstrating effectiveness in the real-world setting, particularly to assess effectiveness, and inform historical controls and address uncertainty. However, the guidance text states that observational data could be used only if IPD are available to allow for rigorous adjustment for confounding. Access to IPD from observational studies and RWE (i.e., registries) may not always be feasible, nor ethical due to data privacy considerations, particularly in rare disease settings. Even if a network meta-analysis or other type of indirect comparison is done using data from RCTs, only variables reported in the studies can be used for adjustment for confounding, or sometimes the only available evidence for a comparator is from a single-arm trial. It would be useful for the methodological guideline to discuss the role and incorporation of real-world evidence (RWE) in JCA. As such, Takeda is concerned that there may be cases where the only available evidence is based on observational studies, particularly when recognising those included in Phase I and II JCA process (orphan diseases, oncology and ATMPs), given the greater potential for EU marketing authorisation on the basis of assessment with non-RCT evidence (i.e., single arm trials). As such, the primary objective should remain the consideration of the available and not exclusively theoretically best-possible evidence. RWE and observational studies may also be a useful source of data to strengthen a network where methods, such as hierarchal models, can be employed.[1] Reference: 1. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			of multiple treatments: combining direct and indirect evidence. BMJ. 2005;331(7521):897-900. doi:10.1136/bmj.331.7521.897	
EFSPI	12	324	Current wording: "However, this requires access to the full individual patient-level data (IPD) information" This is not accurate. One example is the use of observational data in the context of threshold crossing (Eichler et al 2016, DOI: 10.1002/cpt.515). Recommended rewording: "However, this typically requires access to the full individual patient-level data (IPD) information".	Thank you for your suggestion. We have amended the text along the suggested lines.
Sebastian Werner vfa	13	358 / II.3.1.1	It should read "…distribution of known (e.g., sex or age) and unknown effect modifiers..."	The words in brackets are examples for typical effect modifiers (independent of whether they are known or unknown).
M. Ermisch – GKV-Spitzenverband	13	355/3.1.1, 3.1.2	While the preceding statements suggest a clear distinction between "(dis)similarity" and "heterogeneity" the explanations in sections 3.1.1 and 3.1.2 lacks (easily comprehensible) conceptual clarity. Is similarity only relevant regarding effect modifiers (line 358)? Is heterogeneity only statistical heterogeneity observed ex-post in meta-analysis? How are, as a whole, "clinical heterogeneity", and "dissimilarity" distinct? Both seem to arise from trial population as well as (other) trial characteristics.	Similarity and homogeneity are concepts that can overlap and not totally mutually exclusive. We have tried to distinguish them as best as we could. Both can be the consequence of differences in patient characteristics and/ or designs. We have added a sentence in the homogeneity section for clarity.
Mihai Rotaru - EFPIA	13	361 / 3.1.1	**Editorial: Similarity assumption** <u>Current wording:</u> "If dissimilarities between studies in study design and/or patient characteristics are observed at a level that is considered substantial, it can be indicative that the fundamental assumption of exchangeability will not hold." <u>Suggested rewording:</u> Add: "If dissimilarities between studies in study design and/or	In this paragraph, we are only stating confounding can be expected if a substantial amount of dissimilarity is found, which is factual. The possibility for correction is described later in the document.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			patient characteristics are observed at a level that is considered substantial, it can be indicative that the fundamental assumption of exchangeability will not hold. **By substantial dissimilarities we mean differences that cannot be corrected or adjusted for.**" [Note: **bold** denotes proposed inclusion] <u>Rationale:</u> EFPIA believes that the term 'substantial' dissimilarities are subjective and therefore requires further clarification in the document. Adjustment methods when applied would work to address the between study heterogeneity, therefore EFPIA recommends specifying dissimilarities which have not been adjusted for.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	13	361 / 3.1.1	<u>Current wording:</u> "If dissimilarities between studies in study design and/or patient characteristics are observed at a level that is considered substantial, it can be indicative that the fundamental assumption of exchangeability will not hold." <u>Suggested rewording:</u> Add: "If dissimilarities between studies in study design and/or patient characteristics are observed at a level that is considered substantial, it can be indicative that the fundamental assumption of exchangeability will not hold. **By substantial dissimilarities we mean differences that cannot be corrected or adjusted for.**" [Note: **bold** denotes proposed inclusion] <u>Rationale:</u> Takeda believes that the term 'substantial' dissimilarities are subjective and therefore requires further clarification in the document. Adjustment methods when applied would work to address the between study heterogeneity, therefore we recommend specifying dissimilarities which have not been adjusted for.	In this paragraph, we are only stating confounding can be expected if a substantial amount of dissimilarity is found, which is factual. The possibility for correction is described later in the document.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
EFSPI	13	355-367	The fundamental assumption of exchangeability is described in terms of similarity of study design and patient characteristics. However, there is no mention of other similarities, such as the fact that therapies compared should have been available for all the patients included in the evidence synthesis. For example, patients included in the studies 10 years ago may not be comparable with patients in the recent studies, because of, for example, stage migration, and some of the therapies being compared in the HTA being not available for those patients. Further guidance on considerations regarding the time during which evidence was generated]should be provided	We agree that this is a relevant situation. However, it is not the scope of the Guideline to describe all special situations.
Richard Birnie, Lumanity	13	Line 338-344	Tests for statistical heterogeneity are often under powered and over interpreted. Suggest softening this language: "If the test indicates heterogeneity, then variations in between-study treatment effects are plausibly not due to chance (i.e., to random error) alone and thus there are likely systematic errors, so it is necessary to look for dissimilarities that explain this statistical heterogeneity. However, tests for statistical heterogeneity can be under powered and careful consideration should be taken when interpretating p-values." The key point is to emphasise the importance of quantification and investigation of heterogeneity rather than over reliance on quite limited hypothesis tests	We do not think the current wording implies careful considerations. In addition, it is usual knowledge in statistics that an underpowered test does not mean the H0 is true.
Mihai Rotaru - EFPIA	13	329 / 3	**Editorial: Similarity v homogeneity** <u>Current wording:</u> "[…] similarity and homogeneity are sometimes used interchangeably. Since there is no common terminology for the concepts that are described in the following sections, it is possible that these concepts are described with a different terminology elsewhere." <u>Suggested rewording:</u> **"We are considering that different concepts are underpinning each of these terms as described in this section."**	We do not think we need to clarify our statement further.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			[note: **bold** denotes proposed inclusion] <u>Rationale:</u> EFPIA proposes that the authors mention here that the difference they consider between similarity and homogeneity is presented in Sections 3.1.1 and 3.1.2 to clarify that they do not use these terms interchangeably.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	13	329 / 3	<u>Current wording:</u> "[…] similarity and homogeneity are sometimes used interchangeably. Since there is no common terminology for the concepts that are described in the following sections, it is possible that these concepts are described with a different terminology elsewhere." <u>Suggested rewording:</u> "**We are considering that different concepts are underpinning each of these terms as described in this section."** [note: **bold** denotes proposed inclusion] <u>Rationale:</u> We request the authors define the difference they consider between similarity and homogeneity which are presented in Sections 3.1.1 and 3.1.2 to clarify that they do not use these terms interchangeably.	We do not think we need to clarify our statement further.
Mihai Rotaru - EFPIA	13	333 / 3	**Editorial: Exchangeability** <u>Current wording:</u> " […] exchangeability is the most fundamental assumption, […]." <u>Suggested rewording:</u> " […] exchangeability is the most fundamental assumption, […]. **The**	Purpose of the practical guideline.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			suitable methods to investigate the fulfilment of similarity, homogeneity and consistency assumptions of relative effects are described in the practical guideline on direct and indirect comparisons." [note: **bold** denotes proposed inclusion] <u>Rationale:</u> The current document is missing guidance on how HTDs can demonstrate or support the assumptions of exchangeability (i.e., similarity, homogeneity and consistency assumptions) in the underlying analysis. This additional guidance would provide clarity on the expectations of JCA assessors and be important to ensure HTD prepares for and presents adequate evidence for the underlying analysis. EFPIA requests that either the current document be updated to specify the evidence required to demonstrate exchangeability or that this be included in the forthcoming practical guideline (D4.3.1).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	13	333 / 3	<u>Current wording:</u> " […] exchangeability is the most fundamental assumption, […]." <u>Suggested rewording:</u> " […] exchangeability is the most fundamental assumption, […]. **The suitable methods to investigate the fulfilment of similarity, homogeneity and consistency assumptions of relative effects are described in the practical guideline on direct and indirect comparisons."** [note: **bold** denotes proposed inclusion] <u>Rationale:</u> The current document is missing guidance on how HTDs can demonstrate or support the assumptions of exchangeability (i.e., similarity, homogeneity and consistency assumptions) in the underlying analysis. This additional guidance would provide clarity on	Purpose of the practical guideline.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			the expectations of JCA assessors and be important to ensure HTD prepares for and presents adequate evidence for the underlying analysis. We requests that either the current document be updated to specify the evidence required to demonstrate exchangeability or that this be included in the forthcoming practical guideline (D4.3.1).	
Sebastian Werner vfa	13	334 / II.3	Instead of "were substituted to another" it should read "were substituted by other patients".	We replaced "to" with "for".
Mihai Rotaru - EFPIA	13	340 / 3	**Editorial: General statistical considerations** Current wording: "Moreover, there is the possibility of testing for statistical heterogeneity." Suggested rewording: "Moreover, there is the possibility of testing for statistical heterogeneity **[34, 63].**" [note: **bold** denotes proposed inclusion] Rationale: EFPIA suggests that the specific tests for heterogeneity are referenced in the document.	See Section 3.1.2; additionally, we refer to the upcoming Practical Guideline: Direct and Indirect Comparisons (deliverable D4.3.1).
Tanja Podkonjak – Takeda Pharmaceuticals International AG	13	340 / 3	**Editorial: General statistical considerations** Current wording: "Moreover, there is the possibility of testing for statistical heterogeneity." Suggested rewording: "Moreover, there is the possibility of testing for statistical heterogeneity **[34, 63].**" [note: **bold** denotes proposed inclusion]	See Section 3.1.2; additionally, we refer to the upcoming Practical Guideline: Direct and Indirect Comparisons (deliverable D4.3.1).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Rationale: We suggests that the specific tests for heterogeneity are referenced in the document.	
Mihai Rotaru - EFPIA	13	347 / 3	**Editorial: General statistical considerations** Current wording: "… should be carefully assessed before and during assessment when undertaking a formal evidence synthesis." Suggested rewording: "… should be carefully ~~assessed~~ **considered** before and during assessment when undertaking a formal evidence synthesis, **for example using tests described in the literature for similarity [40, 73] and heterogeneity [34, 63].**" " [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] Rationale: EFPIA recommend the proposed revision as we believe the current wording 'assessed before and during assessment' may cause confusion. Additionally, EFPIA suggests that the recommended assessment methods be referenced in the document or linked to subsequent sections.	We deleted the words "and during assessment".
Tanja Podkonjak – Takeda Pharmaceuticals International AG	13	347 / 3	**Editorial: General statistical considerations** Current wording: "… should be carefully assessed before and during assessment when undertaking a formal evidence synthesis." Suggested rewording: "… should be carefully ~~assessed~~ **considered** before and during assessment when undertaking a formal evidence synthesis, **for example using tests described in the literature for similarity [40, 73] and heterogeneity [34, 63].**" "	We deleted the words "and during assessment".

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			[note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] Rationale: We believe the current wording 'assessed before and during assessment' may cause confusion. Additionally, Takeda suggests that the recommended assessment methods be referenced in the document or linked to subsequent sections.	
EFSPI	13	333	Current wording: " […] exchangeability is the most fundamental assumption, […].": We suggest to include guidance on how to assess whether the exchangeability assumption holds in the underlying analysis and which sensitivity analyses could be performed to assess whether the results are robust against violation of this assumption.	Methods to assess the assumptions are described in upcoming Practical Guideline: Direct and Indirect Comparisons (deliverable D4.3.1)
GSK	13	334	Instead of "were substituted to another" it should read "were substituted by other patients".	We replaced "to" with "for".
GSK	13	357-8	Not clear how one can assess the differences in unknown effect modifiers, Recommend more clarity.	More guidance for assessors and co-assessors will be provided in the practical guideline.
GSK	13	358	It should read "..distribution of known (e.g., sex or age) and unknown effect modifiers.."	The words in brackets are examples for typical effect modifiers (independent of whether they are known or unknown).
Sebastian Werner vfa	14	398-400 / II.3.1.3	"Results for sensitivity analyses should be thoroughly discussed in the context of the evidence available and the results obtained." Sensitivity analyses are an adequate tool to investigate and discuss the robustness of results especially in complex situations or in case of multiple sources with different uncertainties. Please consider describing a more detailed approach on what are the advantages and how to handle sensitivity analyses.	We refer to the Practical Guideline D4.5.1: Applicability of Evidence.
Mihai Rotaru - EFPIA	14	364 / 3.1.1	**Similarity Assumption**	We do think this statement is factual in terms of methodological rigor.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Current wording: "Therefore, only if study design and the patient populations are considered similar enough can the results be of value in decision-making." Suggested rewording: "~~Therefore, only if study design and the patient populations are considered similar enough can the results be of value in decision-making~~." Rationale: EFPIA suggests that this sentence be removed as it may be misleading and is not fully accurate. EFPIA agrees that heterogeneity between studies may introduce bias, but if the likely direction of bias is known, a decision may be based on this knowledge. Additionally, EFPIA believes the current text imposes a value judgement and is therefore not in accordance with the HTA regulation.	
Mihai Rotaru - EFPIA	14	366 / 3.1.1	**Editorial: Practical Guideline** Current wording: "Specific guidance on assessing similarity in the JCA is provided in EUnetHTA 21 Practical Guideline D4.3.1 Direct and Indirect Comparisons." Suggested wording: Please correct title or indicate that this is an upcoming publication. Rationale: EFPIA believes this is a typographical error and the reference is incorrect as it seems to refer to the upcoming practical guideline (D4.3.1) yet lists the current guideline title, "Direct and indirect comparisons".	This is not a typographical error; the titles of the Methodological Guidelne (D4.3.2) and the forthcoming Practical Guideline (D4.3.1) are the same.
Mihai Rotaru - EFPIA	14	379 / 3.1.2	**Editorial: Homogeneity**	Thanks. We have amended the text accordingly.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Current wording: "Clinical heterogeneity includes variability in patient inclusion criteria (e.g., age, severity of disease, duration of follow-up), interventions (e.g.,] dosage, administration route) and outcomes (e.g., different time points)." Suggested rewording: "Clinical heterogeneity includes variability in patient inclusion criteria (e.g., age, severity of disease, duration of follow-up), interventions (e.g.,] dosage, administration route) and outcomes (e.g., different time points). **It should be noted that if these differences are effect-modifiers, they should be considered in terms of (dis)similarities.**" [Note: **bold** denotes proposed inclusion] Rationale: EFPIA recommends adding clarifying text on the types of listed differences per suggested text.	
Mihai Rotaru - EFPIA	14	389 / 3.1.2	**Editorial: Heterogeneity** Current wording: "Regardless of whether heterogeneity can be explained there must still be a decision whether or not to proceed with the comparison and whether subgroup analyses will sufficiently explore the impact of the heterogeneity on the analysis outputs." Suggested rewording: "Regardless of whether **between-trial** heterogeneity can be explained there must still be a decision whether or not to proceed with the comparison and whether subgroup analyses will sufficiently explore the impact of the heterogeneity on the analysis outputs." [note: **bold** denotes proposed inclusion]	Thanks. We have amended the text accordingly.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Rationale: EFPIA recommends that a clear distinction should be made in the methodological guideline on subgroup analyses, to assess relative treatment effect amongst relevant patient subgroups, versus indirect comparison using a subset of trials that are considered sufficiently similar to generate reliable estimates of relative treatment effect.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	14	364 / 3.1.1	Current wording: "Therefore, only if study design and the patient populations are considered similar enough can the results be of value in decision-making." Suggested rewording: "~~Therefore, only if study design and the patient populations are considered similar enough can the results be of value in decision-making~~." Rationale: Takeda suggests that this sentence be removed as it may be misleading and is not fully accurate. Takeda agrees that heterogeneity between studies may introduce bias, but if the likely direction of bias is known, a decision may be based on this knowledge. Additionally, the current text imposes a value judgement and is therefore not in accordance with the HTA regulation.	We do think this statement is factual in terms of methodological rigor.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	14	366 / 3.1.1	Current wording: "Specific guidance on assessing similarity in the JCA is provided in EUnetHTA 21 Practical Guideline D4.3.1 Direct and Indirect Comparisons." Suggested wording: Please correct title or indicate that this is an upcoming publication. Rationale: This seems to be a typographical error and the reference is incorrect	This is not a typographical error; the titles of the Methodological Guidelne (D4.3.2) and the forthcoming Practical Guideline (D4.3.1) are the same.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			as it seems to refer to the upcoming practical guideline (D4.3.1) yet lists the current guideline title, "Direct and indirect comparisons".	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	14	379 / 3.1.2	Current wording: "Clinical heterogeneity includes variability in patient inclusion criteria (e.g., age, severity of disease, duration of follow-up), interventions (e.g.,] dosage, administration route) and outcomes (e.g., different time points)." Suggested rewording: "Clinical heterogeneity includes variability in patient inclusion criteria (e.g., age, severity of disease, duration of follow-up), interventions (e.g.,] dosage, administration route) and outcomes (e.g., different time points). **It should be noted that if these differences are effect-modifiers, they should be considered in terms of (dis)similarities.**" [Note: **bold** denotes proposed inclusion] Rationale: Takeda recommends adding clarifying text on the types of listed differences per suggested text.	Thanks. We have amended the text accordingly.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	14	389 / 3.1.2	**Editorial: Heterogeneity** Current wording: "Regardless of whether heterogeneity can be explained there must still be a decision whether or not to proceed with the comparison and whether subgroup analyses will sufficiently explore the impact of the heterogeneity on the analysis outputs." Suggested rewording: "Regardless of whether **between-trial** heterogeneity can be explained there must still be a decision whether or not to proceed with the comparison and whether subgroup analyses will sufficiently explore the impact of the heterogeneity on the analysis outputs."	Thanks. We have amended the text accordingly.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			[note: **bold** denotes proposed inclusion] <u>Rationale:</u> We recommend that a clear distinction should be made in the methodological guideline on subgroup analyses, to assess relative treatment effect amongst relevant patient subgroups, versus indirect comparison using a subset of trials that are considered sufficiently similar to generate reliable estimates of relative treatment effect.	
Liebenhoff, BAH	14	398 - 400	"Results for sensitivity analyses should be thoroughly discussed in the context of the evidence available and the results obtained." Sensitivity analyses are used to investigate the robustness of results, especially in complex situations or in the case of multiple sources with different uncertainties. Therefore, a more detailed approach to the benefits and handling of sensitivity analyses is needed.	We refer to the Practical Guideline D4.5.1: Applicability of Evidence.
GSK	14	381-382	Methodological heterogeneity cannot always be properly assessed as it relies on availability of documents on the study design and SAP. For example, the derivation of outcomes can be assumed but sometimes cannot be confirmed as access to SAP is not achieved.	The lack of availability of the SAP to the HTD is a concern however it does not impact on the guidance for methodological rigour for the assessment of methodological heterogeneity.
James Ryan AZ	14	375	"does not prove homogeneity .. owing to lack of power" Suggest rewording: However, non significance of a statistical test for heterogeneity **does not automatically lead to acceptance of** homogeneity.	We do think our statement is factual. Not rejecting H0 does not mean H0 is true.
James Ryan AZ	14	464	"When the number of studies included is small, random effects methods may have low statistical power". Interpreting statistical power under random effects can be challenging and it might be useful to include recommendations in defining power. See	This article confirms that in general at least 5 studies are needed for a random-effects model. It is out of the scope of this Guideline to describe more technical details.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/jrsm.1240 for evidence and possible solution.	
Sebastian Werner vfa	15	413 / II.3.1.3	Please delete ")".	Thanks, done.
Sebastian Werner vfa	15	437 / II.3.2	What is meant specifically by "… has to provide all available data"? Please consider describing more in detail.	It means all available evidence that can be relevant for the JCA must be provided by the HTD.
Mihai Rotaru - EFPIA	15	414 / 3.2	**Editorial: Sources of bias** EFPIA believes that Section 3.2 regarding the assessment of bias misses some recommendations which should be included. For example, for population adjustment approaches in unanchored cases, it is recommended to provide evidence on the likely extent of error due to unaccounted covariate. Further expanding the recommendations on the assessment of bias is requested.	Thank you – this is issue is covered under section 6 and therefore we don't propose to add it again under this section.
Bayer	15	401-402/ Section 3.1.3	For assessing the influence by one or a small number of studies, the inspection of the h matrix of the respective model has shown to be helpful as discussed here: • *König et al.: Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Statistics in Medicine* 2013,32 5414–5429. *https://onlinelibrary.wiley.com/doi/epdf/10.1002/sim.6001*	The Guideline is not a statistical textbook; it is out of the scope of the Guideline to describe all available methods.
Bayer	15	410-413/ Section 3.1.3	An analysis conducted with and without a particular study to determine the impact of that study shows the result as it would have been without this study, however the contribution of the other studies in the network may also change and especially in network meta-analysis, this can change the connectivity, or the evidence flow in the network, or the possibility to estimate some of the treatment comparisons. Further approaches, e.g., node-splitting approach, was proposed comparing a model where the consistency assumption is	The node-splitting method is described in the upcoming Practical Guideline: Direct and Indirect Comparisons (deliverable D4.3.1).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			relaxed for one treatment comparison to the model assuming consistency across the entire network to highlight inconsistent treatment comparisons within the network. *Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Statist Med.2010;29(7-8):932-944.*	
Richard Birnie, Lumanity	15	Line 424-440	Good guidance on how to detect publication bias is provided. However, further guidance is required on what to do when publication bias is identified. Especially given "publication bias typically only arises in the case of studies sponsored or conducted by other organisations" which is outside the control of the developer of the health technology being assessed.	More guidance for assessors will be provided in the practical guideline. Regarding the last sentence, it is only a factual assessment about the probable source of publication bias.
GSK	15	410-413	Sensitivity can help identify concerns but often there is a trade-off between following the prespecified way of analyses and the data driven direction which may point to the exclusion of a study etc.	We agree, but this has no consequences for this Guideline.
Silke Walleser Autiero Medtronic	15	407-409	These are analogous methods for assessing the assumption of normality of study effects. Non-normality is a real problem and incorrectly assuming normality can introduce a lot of bias, inflate type 1 error etc. with meta-analysis.	The Guideline is not a statistical textbook; it is out of the scope of the Guideline to describe all available methods.
EFSPI	15	417-418	Current wording: "When performing evidence synthesis, there are two potential sources of bias that should be considered" This sentence only mentions internal bias and publication bias. However, external validity is also important in HTA and should be mentioned (indeed, external validity is explicitly mentioned in the HTA regulation, L458/2, Introduction, par 28). Recommended wording: "When performing evidence synthesis, all sources of potential bias should be considered including publication bias and external validity"	The basic assumption of the Guideline is that the available evidence is relevant for the research question (see Introduction). The assessment of external validity is out of the scope of this Guideline.
GSK	15	413	Please delete ")".	Thanks, done.
GSK	15	437	What is meant specifically by "... has to provide all available data"?	It means all available evidence that can be relevant for the JCA must be provided by the HTD.
Mihai Rotaru -	16	472 / 3.4	**Editorial: Frequentist and Bayesian approaches**	The corresponding statement is given in

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
EFPIA			Current wording: to statement in key point 1: "Bayesian methods are particularly useful in situations with sparse data". Suggested rewording: "Bayesian methods are particularly useful in situations with sparse data **where the fixed-effects assumption is not adequate.**" [note: **bold** denotes proposed inclusion] Rationale: EFPIA recommends adding text to clarify when Bayesian methods may be useful.	the Key Points. Nevertheless, the more general statement in the text is correct.
Mihai Rotaru - EFPIA	16	477 / 3.4	**Editorial: Frequentist and Bayesian approaches** Current wording: "When noninformative prior distributions are used, results are frequently equivalent to those observed using a frequentist approach". Suggested wording: "When non-informative prior distributions are used, results ~~are frequently equivalent~~ **can be similar** to those observed using a frequentist approach." [note: **bold** and ~~strikethrough~~ denotes proposed inclusion and deletion, respectively] Rationale: EFPIA believes the statement is too strong and absolute and recommends revising and softening the language in the document as suggested.	Thank you. We have amended the text accordingly.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Bayer	16	466-467/ Section 3.3	Current: or a qualitative summary of the study results might also be considered. Proposed: delete Rationale: this is not an evidence synthesis approach.	We acknowledge that a qualitative summary of study results is not always regarded as an evidence synthesis approach. However, in the context described we still believe that it may provide useful information for HTA, particularly in situations where formal evidence synthesis is not feasible. Therefore, we have not deleted the highlighted text.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	16	472 / 3.4	<u>Current wording:</u> to statement in key point 1: "Bayesian methods are particularly useful in situations with sparse data". <u>Suggested rewording:</u> "Bayesian methods are particularly useful in situations with sparse data **where the fixed-effects assumption is not adequate.**" [note: **bold** denotes proposed inclusion] <u>Rationale:</u> Takeda recommends adding text to clarify when Bayesian methods may be useful.	We think the general statement does not need to be restricted.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	16	477 / 3.4	<u>Current wording:</u> "When noninformative prior distributions are used, results are frequently equivalent to those observed using a frequentist approach". <u>Suggested wording:</u> "When non-informative prior distributions are used, results ~~are~~	Thank you. We have amended the text accordingly.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			~~frequently equivalent~~ **can be similar** to those observed using a frequentist approach." [note: **bold** and ~~strikethrough~~ denotes proposed inclusion and deletion, respectively] <u>Rationale:</u> Takeda believes the statement is too strong and absolute and recommends revising and softening the language in the document as suggested.	
Sarah Smith, Lumanity	16	Lines 453-458	The text here reads: "Without adequate justification that the assumption of a common effect holds, a random-effects model should generally be used. There are situations in which a fixed-effect model is appropriate, such as a pairwise meta-analysis of two studies with identical designs." It is not clear what an "adequate justification" would entail. Is it possible to provide more guidance? Equally, it is not clear in what other situations except for a "pairwise meta-analysis of two studies with identical designs" that a fixed-effects analysis would be appropriate. Please could more guidance be provided to explain in which other situations a fixed-effects analysis is justified.	Situations in which a fixed-effect model can be acceptable are described above. Nonetheless, we have changed the structure of the paragraph for more clarity.
Richard Birnie, Lumanity	16	Lines 464-468	The text here reads "When the number of studies included is small, random-effects methods may have low statistical power. In this scenario, a fixed-effect approach can be an option if appropriate, or a qualitative summary of the study results might also be considered [3,67]. Bayesian methods are also an option in cases involving sparse data and few studies [3]." This text is ambiguous and requires clarification to avoid misuse. Our interpretation here is that if there is insufficient data to reliably	Thank you. We have amended the draft for more clarity

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			estimate the random effect then a fixed effect model could be an option if the assumptions hold. However, text above (lines 444-446) indicates the assumptions will rarely hold, if ever. It follows that a Bayesian approach using informative priors to support estimation of the random effect will be the most appropriate method in many cases and the guidance should be amended to state this more clearly.	
Silke Walleser Autiero Medtronic	16	453-454	Agreed. No mention of the common assumption of normality though. These models are *not* robust to non-normality with a small to moderate number of studies. In fact, they reduce to a t-test if the within-study sample sizes are large. t-tests are well-known to suffer from bias and inflated type I error in small samples	While we agree with the comment, we think that this is too technical to be included in this Guideline.
James Ryan AZ	16	477	"Prior distributions that have broad support in the parameter space are called uninformative". Suggestion "Prior distributions that have broad support in the parameter space are called **vague priors**". Rationale: It may be argued that all priors are to some extent informative	Thank you. While this is indeed a good point, "noninformative" is still frequently used. Therefore, we have proposed the two terminologies within the draft.
Sebastian Werner vfa	16	479	Please change "are frequently equivalent to those observed using a frequentist approach "to "are frequently equivalent to those observed using a frequentist approach albeit usually larger variability". Credible interval from Bayesian tends to be wider in many cases.	We do not think it is possible to make such a general statement.
Sebastian Werner vfa	17	495-498 / III.3.5	In section 3.5 hybrid methods to combine IPD and aggregated study data are mentioned. This should be supplemented by a hybrid method that combines reconstructed individual patient data e.g., by digitization technics for time-to event data from published KM curves. So, we suggest to change the sentence to: "Hybrid methods	Thank you for your comment. We have added in a reference to the Guyot paper and a reference to the Di Pietrantonj (2006) in this section, as well as some of the suggested text.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			are also available for combining IPD and aggregated study data [59]. Technics to reconstruct individual patient data from published time-to-event data (Kaplan-Meier plot) could be considered also (Ref.: Guyot, P., Ades, A., Ouwens, M.J. et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 12, 9 (2012). https://doi.org/10.1186/1471-2288-12-9) to provide better estimates for relative effectiveness. Evidence synthesis based on IPD or reconstructed IPD have better modelling options for estimating treatment effectiveness when compared to corresponding aggregate data analyses."	
Mihai Rotaru - EFPIA	17	502 / Key Points I	**Verification of assumptions for consistency, similarity, and homogeneity** Current wording: "Sufficient similarity and sufficient homogeneity are required to justify an evidence synthesis of the data being considered." Suggested rewording: EFPIA recommends wording such as "**Although the different analytic approaches rely on key assumptions such as consistency, similarity and homogeneity, these are not always verifiable in an unambiguous way**", is added in the Summary and Section 3 of the document. [Note: **bold** denotes proposed inclusion] Rationale: EFPIA recommends that it should be made explicit that some assumptions made when undertaking an analysis may be difficult to verify in practice when undertaking evidence synthesis for the purposes of a JCA.	We think that our statement is correct as it is. Methods to assess these assumptions are described in upcoming Practical Guideline: Direct and Indirect Comparisons (deliverable D4.3.1).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Bayer	17	494-495 / Section 3.5	Section 3.5 explains that IPD can be analysed in 2 ways (one and two step approach) – the advantage of the two step approach is not very clear here as the first step is to analyse the studies separately which would not require to have IPD.	We do not understand the comment. The first step (separate analysis of the studies) IPD are required.
EFSPI	17	Key points I	Recommend to add a bullet point: "Discussion of robustness of results as assessed via sensitivity analyses".	Thank you; we added this to the Key Points I.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	17	502 / Key Points I	**Verification of assumptions for consistency, similarity, and homogeneity** Current wording: "Sufficient similarity and sufficient homogeneity are required to justify an evidence synthesis of the data being considered." Suggested rewording: Takeda recommends wording such as "**Although the different analytic approaches rely on key assumptions such as consistency, similarity and homogeneity, these are not always verifiable in an unambiguous way**", is added in the Summary and Section 3 of the document. [Note: **bold** denotes proposed inclusion] Rationale: It should be made explicit that some assumptions made when undertaking an analysis may be difficult to verify in practice when undertaking evidence synthesis for the purposes of a JCA.	We think that our statement is correct as it is. Methods to assess these assumptions are described in upcoming Practical Guideline: Direct and Indirect Comparisons (deliverable D4.3.1).
EFSPI	17	495-498	Current wording: "Hybrid methods are also available for combining IPD and aggregated study data" To properly reflect methods most commonly used, this should be supplemented by a hybrid method that combines reconstructed	Thank you for your comment. We have added in a reference to the Guyot paper and a reference to the Di Pietrantonj (2006) in this section, as well as some of the suggested text.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			individual patient data e.g., by digitization technics for time-to event data from published KM curves. Recommended rewording: "Hybrid methods are also available for combining IPD and aggregated study data [59]. Techniques to reconstruct individual patient data from published time-to-event data (Kaplan-Meier plot) could be considered also (Ref.: Guyot, P., Ades, A., Ouwens, M.J. et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 12, 9 (2012). https://doi.org/10.1186/1471-2288-12-9) to provide better estimates for relative effectiveness. Evidence synthesis based on IPD or reconstructed IPD have better modelling options for estimating treatment effectiveness when compared to corresponding aggregate data analyses."	
GSK	17	490-491	Suggestion for additional wording (additions shown in capitals):"preferred to statistical analyses that use only summary statistics BECAUSE THE DISTRIBUTION CAN BE USED TO DERIVE HIGHER-ORDER MOMENTS, SUBGROUP ANALYSES AND ASSESS DISTRIBUTIONAL ASSUMPTIONS".	Thank you. We have amended the draft
Silke Walleser Autiero Medtronic	17	489-490	Also typically assume/We recommend to assume exchangeability. However, it is easier to test this assumption with IPD by including a study treatment interaction.	We already described that evidence syntheses based upon IPD have better modelling options. However, a test of the study-treatment interaction is not an "easy test of the exchangeability assumption".
Silke Walleser Autiero Medtronic	17	511-512	Sensitivity analyses should be performed, especially for the prior on the variance of the underlying treatment effects. There is no such thing as 'non-informative' here.	We agree; this is described in Section 3.4; not all important points can be included in the key point boxes.
Richard Birnie, Lumanity	17	Lines 483/484	There are some cases where an informative prior on a treatment effect may be appropriate. For example, the use of non-randomised evidence may provide some prior information on the relative treatment effect in a (network) meta-analysis of RCTs. Particularly for older comparator treatments that may have been used in clinical practice for some time. The wording "informative prior distributions	Thank you for your comment; however, in the HTA applications the current wording is adequate; rare exceptional cases are covered by the word "generally".

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			should generally only be used for the heterogeneity parameter and not for the treatment effect itself" does not allow for this scenario. Suggest changing the wording to: "informative prior distributions should generally only be used for the heterogeneity parameter unless there is credible justification for an informative prior on the treatment effect itself."	
Matias Olsen, EUCOPE	17	500	Would a fractional polynomial analysis based on reconstructed patient level data be an acceptable approach when IPD is not available (when proportional hazard assumption is not violated)?	Yes, provided that the reconstruction was performed adequately.
M. Ermisch – GKV-Spitzenverband	17	501	The insertion of a "key points"-box in section 3 seems superfluous. The section deals with various more or less cross-cutting and diverse issues. Highlighting them in a box runs the risk to over-simplify these. In addition, mentioning particular techniques like e.g. meta-regression may give rise to the expectation that a more elaborate discussion/explanation can be found in the text of section 3 – which is not the case in this example.	We disagree and think that the Key Points I are useful.
EFSPI	17	504	Current wording: "If heterogeneity is too strong to justify an evidence synthesis" Recommend to include a clarification what is meant by "heterogeneity too strong"	We refer to the Practical Guideline (deliverable 4.3.1)
Bayer	18	553-554 / Section 4.1	Current: The standard method for random-effects meta-analyses is the Knapp-Hartung (KH) method, also the called Hartung-Knapp-Sidik-Jonkmann method. Proposed: An accepted alternative method to the established DerSimonian and Laird is the Knapp-Hartung method. Rationale: This is one of the accepted methods next to the established DerSimonian and Laird REF MA.	We disagree; the DerSimonian-Laird-method is no longer accepted; please read Section 4.1 and the corresponding references.
Matias Olsen, EUCOPE	18	550-553	The method of Knapp-Hartung is only suitable if more than 5 studies are available, which is usually not the case for new products. The most systematic reviews, including Cochrane reviews which are also the basis for the determination of the comparators, use DerSimonian and	We disagree; the DerSimonian-Laird-method is no longer accepted; please read Section 4.1 and the corresponding references.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Laird. This is also the method provided in the ReviewManager which is used for the production of Cochrane reviews. Therefore, it should not be excluded. Both Knapp-Hartung as well as the Bayesian models are not sufficiently reliable and therefore not appropriate for <5 studies. A qualitative interpretation cannot replace the statistical approach of DerSimonian and Laird. The whole chapter 4.1 would need to be adjusted accordingly. Replace: "The most common estimation method for the random-effects model ~~was~~ the method of DerSimonian and Laird. ~~However, this method has increased type 1 errors (i.e., p-values that are too small and confidence intervals that are too narrow), especially in the case of few available studies, and is no longer recommended."~~ With: "The most common estimation method for the random effects model **is** the method of DerSimonian and Laird. **It is commonly used and recommended in the case of few (<5) available studies.**".	It is planned to implement Knapp-Hertung and other adequate methods for random-effects meta-analyses in RevMan, see https://methods.cochrane.org/sites/default/files/public/uploads/statistics_mg_summary_report_2020_0.pdf
James Ryan AZ	18	550-553	The most common method … was the DerSimonian and Laird … is no longer recommended". A single reference [14] is used to support this conclusion about this standard method. The authors may also want to consider and reference Langan et al. (2018) - https://doi.org/10.1002/jrsm.1316	This statement is also supported by the references given in the next sentences. We added some of these to the statement in line 553.
Silke Walleser Autiero Medtronic	18	551-553	Important point about the DerSimonian and Laird method. Please leave in.	See the replies to the comments regarding lines 550-553 above
Liebenhoff, BAH	18	526 - 527	"The summary statistic can be an odds ratio, risk ratio, risk difference, hazard ratio, difference of means or standardised mean	The formulation "can be" means that there are also other possibilities

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			difference." The measures quoted above are just a selection of appropriate measures. For that, please consider adding "e.g.".	(although this list is nearly complete).
Sebastian Werner vfa	18	526-527 / II.4	"The summary statistic can be an odds ratio, risk ratio, risk difference, hazard ratio, difference of means or standardised mean difference." There are several more measures. Please consider adding 'e.g.'.	The formulation "can be" means that there are also other possibilities (although this list is nearly complete).
James Ryan AZ	18	515, Section 4 Direct Comparisons	**Direct comparisons and inconsistency** A series of separate direct comparisons can result in results that are inconsistent across outcomes because different studies contribute to different analyses. NMA is often considered preferable simply on the grounds that it allows all data to contribute and provide inferences that are not inconsistent in this way. This section would benefit from this consideration.	Section 4 deals with direct comparisons only; indirect comparisons and NMA are described in Section 5.
EFSPI	19	560-566	Several methods are mentioned if few studies are available (<5). We recommend to clarify that application and argumentation for the use of one particular method, accompanied by appropriate sensitivity analyses, will be sufficient.	The document does not imply all methods have to be performed.
James Ryan AZ	19	554-559	The authors may want to consider a discussion on the KH method, including critiques and possible modifications. https://doi.org/10.1002/sim.7411 and https://doi.org/10.1002/sim.6879 and https://doi.org/10.1002/jrsm.1356	We agree; these modifications are discussed in the Practical Guideline (deliverable 4.3.1).
James Ryan AZ	19	567-572	The authors may want to consider https://doi.org/10.1002/sim.7588 which compares seven random-effects models for meta-analyses that estimate the summary odds ratio	Thank you for this hint; however, not all interesting articles can be cited in the Guideline.
EFSPI	19	567-572	For studies with zero-events, it is recommended to use a continuity correction, which is not a robust method. An alternative suggested in	We added the GLMM approach especially for the case of sparse data.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			the document is to use beta-binomial method. Suggest to recommend GLMM approaches. In general, one-step methods should be considered in the presence of small total event counts or sample sizes and very low or high event rates.	
Silke Walleser Autiero Medtronic)	19	557-558	Agree that prediction intervals are important but would be good to state why this is.	Thank you. We have clarified this aspect within the draft.
EFSPI	20	Key points II	We suggest to include a recommendation for time-to-event data	We do not see the need to recommend in general the use of time-to-event data.
Mihai Rotaru - EFPIA	20	Key Points II, Bullet 1	**Editorial: Frequentist methods key points** EFPIA recommends the Peto method is also included in the key point summary for completeness in the document.	Due to the limitations of the Peto method (see Section 4.1), this method should not be included in the Key Points.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	20	Key Points II, Bullet 1	**Editorial: Frequentist methods key points** Takeda recommends the Peto method is also included in the key point summary for completeness in the document. This is stated in the main text 4.1, and we recommend it also be included in the Key Points summary for completeness.	Due to the limitations of the Peto method (see Section 4.1), this method should not be included in the Key Points.
M. Ermisch – GKV-Spitzenverband	20	585	See above regarding the necessity of key points boxes.	We disagree and think that the Key Points are useful.
Sebastian Werner vfa	21	588 – 589 / II.5	It should read "via a common comparator or common comparators" instead of "common comparators"	We rephrased the sentence.
Richard Birnie, Lumanity	21	Lines 608-611	It is agreed that a thorough assessment of the assumptions is required. However, realistically, few analyses will meet the entirety of these requirements and dismissing these analyses as 'not providing a meaningful estimate' when any single assumption is not perfectly met will often mean no evidence at all is available to decision makers. This is not helpful to decision makers and a more	See general response

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			nuanced approach that assesses the impact of an invalidated assumption on the treatment effect would be more pragmatic. It is not the case that such analyses provide no information at all. They do tell us something. Although we fully agree that considerable care is required in the interpretation of any analysis. We suggest, instead, including more guidance on assessing the impact of methodological assumptions on the results through sensitivity analysis.	
M. Ermisch – GKV-Spitzenverband	21	609-611	Indirect adjusted comparisons should only be done if inevitable, e.g. if there are no RCTs for the relevant comparison or if the simultaneous comparison of several interventions is the PICO. Valid conclusions can only be drawn if all transitivity-control measures indicated so (i.e. absence of dissimilarities, absence of heterogeneity and absence of inconsistent effect estimates of direct and indirect effect components). If there are indications of violations of transitivity in the network, e.g. due to a heterogeneous meta-analysis of one direct comparison, plausibly indicative of an effect modifier, the PICO should be separated into the appropriate number of PICOs.	This is described in Section 3.1.2 on heterogeneity and with more details in the Practical Guideline.
EFSPI	21	609-611	Current wording: "If at least one of these requirements is not fulfilled, the results of an adjusted indirect comparison do not provide a meaningful estimate of the treatment effect." Wording suggests that it is either/or which is not reflective of the fact that model fit is assessed as a continuum. Recommended rewording: "if there is evidence that at least one of these assumptions is not fulfilled, it raises concerns about potential bias of the adjusted indirect comparison"	With "not fulfilled" we don't mean a slight deviation, we mean a strong, relevant deviation. The result is not simply a bias. The result is that it is unclear what is estimated with the corresponding analysis. We believe the proposed statement does not provide meaningful information to assessors. Nonetheless, we do agree that the acceptability of an NMA is on spectrum rather than binary. Thus, the wording to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
GSK	21	588 - 589	It should read "via a common comparator or common comparators" instead of "common comparators"	We rephrased the sentence.
M. Ermisch – GKV-Spitzenverband	21	587/5	See comment on p. 9 line 261: A more nuanced statement seems preferable like "… might be used, if an effectiveness estimate is sought". It may happen that no meaningful treatment effect estimate can be provided (line 610, and in Key Points III).	Thanks. We have amended the draft for clarity.
Mihai Rotaru - EFPIA	21	595 / 5 Repeated 26; 764-765	**ITC – Comparisons with nonrandomised evidence and access to IPD** <u>Current wording:</u> "Comparisons based on nonrandomised evidence require access to full IPD information." <u>Suggested rewording:</u> "Comparisons based on nonrandomised evidence require access ~~to full IPD information~~ **to full IPD for at least one trial**." [Note: **bold** and strikethrough denotes proposed inclusion and deletion, respectively] <u>Rationale:</u> EFPIA wishes to highlight that HTDs will seldom, if ever, have access to IPD data for all the trials in a network. While this is the ideal situation, population adjusted methods do not require full IPD information as they can be undertaken if IPD is available for at least one trial. This requirement will be met as the HTDs will have IPD for their trial(s). As such, EFPIA requests the amendment to this current statement.	An analysis without IPD does not have better reliability, if IPD are not available.
Tanja Podkonjak – Takeda Pharmaceuticals	21	595 / 5 Repeated 26; 764-765	**ITC – Comparisons with nonrandomised evidence and access to IPD** <u>Current wording:</u> "Comparisons based on nonrandomised evidence require access to	An analysis without IPD does not have better reliability, if IPD are not available.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
International AG			full IPD information." Suggested rewording: "Comparisons based on nonrandomised evidence require access ~~to full IPD information~~ **to full IPD for at least one study**." [Note: **bold** and strikethrough denotes proposed inclusion and deletion, respectively] Rationale: HTDs will seldom, if ever, have access to IPD data for all the trials in a network. While this is the ideal situation, population adjusted methods do not require full IPD information as they can be undertaken if IPD is available for at least one trial. This requirement will be met as the HTDs will have IPD for their trial(s). As such, Takeda requests the amendment to this current statement.	
Mihai Rotaru - EFPIA	21	597 / 5	**Editorial: Indirect comparisons** Current wording: "The use of methods for indirect comparison based on aggregated data is not recommended in disconnect networks" Suggested rewording: "The use of methods for indirect comparison based **entirely** on aggregated data is not recommended in disconnect networks." [note: **bold** denotes proposed inclusion] Rationale: In many instances, IPD data for all comparators will not be available to the HTD. IPD data for comparators may not be available for competitor data due to commercial sensitivity. Likewise, IPD data from RWE or observational studies may not be available due to accessibility and patient data privacy. However, if IPD are available	An analysis without IPD does not have better reliability, if IPD are not available.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			for one of the interventions (likely the one undergoing JCA by the HTD), it is possible to conduct an NMA in a disconnected network through population adjustment techniques on the intervention where IPD are available. EFPIA recommend the sentence be revised to reflect disconnected networks where no IPD and only aggregate data across all treatments are available as here population adjusted methods cannot be conducted.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	21	597 / 5	<u>Current wording:</u> "The use of methods for indirect comparison based on aggregated data is not recommended in disconnect networks" <u>Suggested rewording:</u> "The use of methods for indirect comparison based **entirely** on aggregated data is not recommended in disconnect networks." [note: **bold** denotes proposed inclusion] <u>Rationale:</u> In many instances, IPD data for all comparators will not be available to the HTD. IPD data for comparators may not be available for competitor data due to commercial sensitivity. Likewise, IPD data from RWE or observational studies may not be available due to accessibility and patient data privacy. However, if IPD are available for one of the interventions (likely the one undergoing JCA by the HTD), it is possible to conduct an NMA in a disconnected network through population adjustment techniques on the intervention where IPD are available. We recommend the sentence be revised to reflect disconnected	An analysis without IPD does not have better reliability, if IPD are not available.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			networks where no IPD and only aggregate data across all treatments are available as here population adjusted methods cannot be conducted.	
Mihai Rotaru - EFPIA	21	609 / 5	**Editorial: Indirect comparisons assumptions** Current wording: "If at least one of these requirements is not fulfilled, the results of an adjusted indirect comparison do not provide a meaningful estimate of the treatment effect." Suggested rewording: "If at least one of these requirements is not fulfilled, the results of an adjusted indirect comparison ~~do not provide a meaningful estimate of the treatment effect~~ **are accompanied by higher uncertainty and the derived treatment effect may be questioned.**" Rationale: EFPIA believes that a violation of an assumption does not automatically lead to a meaningless result as suggested in the current statement.	We believe the proposed statement does not provide meaningful information to assessors. Nonetheless, we do agree that the acceptability of an NMA is on spectrum rather than binary. Thus, we have altered the wording to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".
Tanja Podkonjak – Takeda Pharmaceuticals International AG	21	609 / 5	Current wording: "If at least one of these requirements is not fulfilled, the results of an adjusted indirect comparison do not provide a meaningful estimate of the treatment effect." Suggested rewording: "If at least one of these requirements is not fulfilled, the results of an adjusted indirect comparison ~~do not provide a meaningful estimate of the treatment effect~~ **are accompanied by higher uncertainty and the derived treatment effect may be questioned.**"	We believe the proposed statement does not provide meaningful information to assessors. Nonetheless, we do agree that the acceptability of an NMA is on spectrum rather than binary. Thus, we have altered the wording to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Rationale: A violation of an assumption does not automatically lead to a meaningless result as suggested in the current statement.	
Roche	21	595-596/5	This sentence suggests that all methods for disconnected networks require IPD for all studies. We recommend updating the text to cover also methods such as unanchored STC and MAIC, which have been developed for the case of disconnected networks where a mix of IPD and AD are available - as the guidance correctly points out elsewhere.	In the document, it is described why these methods should be avoided in that context.
Bayer	21	609-611/ Section 5	Current: If at least one of these requirements is not fulfilled, the results of an adjusted indirect comparison do not provide a meaningful estimate of the treatment effect. Proposed: If at least one of these requirements is not fulfilled, the results of an adjusted indirect comparison are accompanied by higher uncertainty and the derived treatment effect may be questioned Rationale: A violation of an assumption does not lead automatically to a useless estimate as suggested in the above statement. Assumptions are not axioms and apodictical within hypothesis logic.	We do agree that the acceptability of an NMA is on spectrum rather than binary. Thus, we have altered the wording to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".
Dr Martin Danner BAG SELBSTHILFE	21	609	After "reported" should be added: "For the goal of the assessment is the determination of an additional benefit or the absence of an additional benefit, the requirements for indirect comparisons are not only a scientific question. In MS, where the result of the assessment has the consequence of reimbursement or not, the requirements might be lower than in MS, where the assessment has only a consequence for the pricing. Then the requirements might be more demanding. Therefore the patient organizations of the MS should be involved in the definition of the requirements."	Thank you for your comment. As this guideline aims to address only the scientific aspects associated with evidence synthesis and not the consequences leading to the reimbursement or otherwise we consider this comment to be outside of the scope of the guideline.
Sebastian Werner vfa	22	Key Points III	"If sufficient similarity, sufficient homogeneity and sufficient consistency cannot be assumed, an adjusted indirect comparison should not be performed because the corresponding results do not	We believe the proposed statement does not provide meaningful information to assessors. Nonetheless, we do agree

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			provide a meaningful estimate of the treatment effect." Change to: "If sufficient similarity, sufficient homogeneity, and sufficient consistency cannot be assumed, sources, magnitude and potential impact on the results has to be discussed. Combining evidence from several sources or sensitivity analyses might be considered. "	that the acceptability of an NMA is on spectrum rather than binary. Thus, we have altered the wording to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".
M. Ermisch – GKV-Spitzenverband	22	616/5/Key Points III	As mentioned, key point boxes seem superfluous. In case of key points III, we advocate for including the remarks in the text of the previous paragraph; introducing new aspects in the box is irritating to the reader.	In fact, Key Points Box 3 refers not only to the paragraph above but also to 5.1 and 5.2. We didn't want to make more boxes with very few key points and applied this approach as compromise.
Liebenhoff, BAH	22	Key Points III	"If sufficient similarity, sufficient homogeneity and sufficient consistency cannot be assumed, an adjusted indirect comparison should not be performed because the corresponding results do not provide a meaningful estimate of the treatment effect." Consider changing to: "If sufficient similarity, sufficient homogeneity and sufficient consistency cannot be assumed, sources, magnitude and potential impact on the results has to be discussed. Combining evidence from several sources or sensitivity analyses might be considered."	We believe the proposed statement does not provide meaningful information to assessors. Nonetheless, we do agree that the acceptability of an NMA is on spectrum rather than binary. Thus, we have altered the wording to read "unlikely to provide a meaningful estimate" rather than "do not provide a meaningful estimate".
Bayer	22	635-637/ Section 5.1	NMA methods that appropriately incorporate random effects are not only available in Bayesian approaches. A reference to an example also using frequentist methods should be given, especially since frequentist and Baysian approaches for NMA are described in different sections in this document.	We added the reference to Section 5.2.1.
EFSPI	22	627-630	The paragraph may be confusing. The word independent is trial specific. If a closed loop is exclusively informed from three-arm trials, then the independence does not hold.	Thank you for your comment. We have expanded this sentence for clarification.
Richard Birnie, Lumanity	22	Line 618-620	It is unclear what a "simple star network" is for a Bucher comparison. Star networks could have more than three treatments, e.g a star with 5 points and one common comparator in the centre (often placebo). It would be possible to do multiple Bucher	Thank you. Text has been amended as suggested.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			comparisons in such a network but an NMA may be preferable. Suggest referring to Figure 1 (a) instead: "Bucher et al. [9] presented an adjusted indirect method of treatment comparison for aggregate data that can estimate relative treatment effectiveness for a simple network which includes three different treatments (Figure 1a)."	
EFSPI	22	key points III/bullet 1	Current wording: "Indirect comparisons are associated with greater uncertainty than direct comparisons." It is unclear what is meant by 'greater uncertainty': a properly conducted indirect comparison between two large studies can be less uncertain (in the sense of providing stronger evidence on the treatment effect) than a direct comparison within a small study. Recommended rewording: "Indirect comparisons are generally at higher risk of bias and uncertainties than direct comparisons". In the second line, Current wording: "[…]direct comparisons should be preferred where possible." May suggest that indirect evidence should not be considered when direct evidence is available. It is recommended to clarify that all relevant evidence is allowed to be taken into account.	Thank you. We have clarified the text accordingly.
Mihai Rotaru - EFPIA	22	Key Points III / Bullet point 3	**Editorial: Assumptions of similarity, homogeneity and consistency** <u>Current wording</u> If sufficient similarity, sufficient homogeneity, and sufficient consistency cannot be assumed, an adjusted indirect comparison should not be performed because the corresponding results do not provide a meaningful estimate of the treatment effect.	We do think that if any of these assumptions cannot be assumed, such comparison should not be performed.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Suggested rewording: "If sufficient similarity, sufficient homogeneity, and sufficient consistency cannot be assumed, an adjusted indirect comparison should ~~not be performed because the corresponding results do not provide a meaningful estimate of the treatment effect~~ **be interpreted with appropriate caution**." [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] Rationale: EFPIA wishes to highlight these key assumptions are not always verifiable in an unambiguous manner. As such, EFPIA recommends the proposed amendment in the document. Furthermore, there may be situations in a JCA in which these may be required particularly due to the potential for multiple PICOs (e.g., alternative comparators).	
EFSPI	22	key points III/bullet 3	Current wording: "If sufficient similarity, sufficient homogeneity and sufficient consistency cannot be assumed, an adjusted indirect comparison should not be performed" If there is no additional evidence is available, it should still be possible to perform an indirect treatment comparison for completeness, with all the caveats in place. Generally, we recommend an inclusive approach towards including all relevant available evidence, accompanied by appropriate caveats and sensitivity analyses to evaluate the robustness of conclusions to limitations in the data and modelling assumptions. Recommended rewording: "[…] an adjusted indirect comparison depends on similarity, homogeneity and consistency, it may be biased and should be interpreted cautiously"	We do think that if any of these assumptions cannot be assumed, such comparison should not be performed.
Tanja Podkonjak –	22	Key Points III	Current wording	We do think that if any of these assumptions cannot be assumed, such

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Takeda Pharmaceuticals International AG		/ Bullet point 3	If sufficient similarity, sufficient homogeneity, and sufficient consistency cannot be assumed, an adjusted indirect comparison should not be performed because the corresponding results do not provide a meaningful estimate of the treatment effect. Suggested rewording: "If sufficient similarity, sufficient homogeneity, and sufficient consistency cannot be assumed, an adjusted indirect comparison should ~~not be performed because the corresponding results do not provide a meaningful estimate of the treatment effect~~ **be interpreted with appropriate caution**." [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] Rationale: Takeda notes that these key assumptions are not always verifiable in an unambiguous manner. As such, we recommend the proposed amendment in the document. Furthermore, there may be situations in a JCA in which these may be required particularly due to the potential for multiple PICOs (e.g., alternative comparators).	comparison should not be performed.
Silke Walleser Autiero Medtronic	22	638	Should ranking of treatments be undertaken in NMA – if yes, which method is preferred? e.g. SUCRA	This depends on the research question. We don't want to give preference to a special method.
James Ryan AZ	23	660-672	The authors state that the Lumley's method "no longer has major practical relevance and is rarely used". The authors may want to consider the following publications and revisit this position. https://link.springer.com/article/10.1186/s12874-016-0184-5 https://doi.org/10.1002/sim.6752 https://doi.org/10.1002/sim.6188	Thank you; however, all 3 articles describe generalisations of the Lumley method underlining our statement that the Lumley method no longer has practical relevance due to its limitations.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
GSK	23	651-659	Sources of inconsistency are harder to identify in a complex network. Should one build a smaller network before extended it further? For example, would one only looking at direct comparisons be the starting point?	We refer to the cited textbooks at the end of the Introduction and the forthcoming Practical Guideline (deliverable 4.3.1).
Silke Walleser Autiero Medtronic	23	644-645	Including all relevant comparators into an NMA is best practice but should be defined *a priori* in the PICO. How will the inclusion of additional comparators (beyond those of interest) be handled?	Even comparators beyond those of primary interest may be relevant, e.g., to connect a network.
EFSPI	23	654	The referenced paper does not support the claim: it discusses covariate adjustment approaches and how this lowers power (since number of studies per treatment identifies regression coefficients)	We agree, thank you; we changed the reference.
EFSPI	23	660	The draft guidance discusses the Lumley method which is rarely used and no longer has major practical relevance. Recommend to add a different reference.	See Sections 5.2.1 and 5.2.2
Mihai Rotaru - EFPIA	24	692 / 5.2.4	**Editorial: Time-varying hazard ratios** Current wording: "Other emerging methods for time-varying hazards ratios described in the literature may also be considered [88]." Suggested rewording: "Other emerging methods for time-varying hazards ratios described in the literature may also be considered [88]**,[89]**" [note: **bold** denotes proposed inclusion] Rationale: EFPIA suggests the addition of a reference (Freeman, 2022)[1] to substantiate this statement. The proposed referenced paper provides details of other methods for undertaking NMA when non-proportional hazards are encountered. These methods include Royston-Parmar cubic spline models, restricted mean survival time, and piecewise exponential models. References: 1. Freeman SC, Cooper NJ, Sutton AJ et al. Challenges of modelling	Thank you for your comment. We have included this reference.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			approaches for network meta-analysis of time-to-event outcomes in the presence of non-proportional hazards to aid decision making: application to a melanoma network. Statistical Methods in Medical Research. 2022 (available at: https://doi.org/10.1177/09622802211070253).	
Mihai Rotaru - EFPIA	24	697 / 5.3	**Editorial: Population-adjusted methods for indirect comparison** Current wording: "When this assumption does not hold, these methods do not yield meaningful results." Suggested rewording: "When this assumption does not hold, ~~these methods do not yield meaningful~~ results **are accompanied by higher uncertainty and the derived treatment effect may be questioned."** [note: **bold** and ~~superscript~~ denotes proposed inclusion and deletion, respectively] Rationale: EFPIA believes that a violation of an assumption does not automatically lead to a meaningless result as suggested in the current statement.	Thank you. We have clarified the text.
Mihai Rotaru - EFPIA	24	703 / 5.3	**Editorial: Relevant effect modifiers** Current wording: "It is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data." Suggested rewording: "It is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan	Thank you. We have clarified the text.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			before analysis of the data. **If this is not given, rationale needs to be provided as to why the identification of one or more relevant effect modifiers has occurred later.**" [note: **bold** denotes proposed inclusion] <u>Rationale:</u> EFPIA recommends the inclusion of the proposed statement into the document for instances where this is not fulfilled, or if new effect modifiers have been identified at a later point. For example, emerging data on a biomarker or effect modifiers identified in recent research with a strong biological plausibility.	
Roche	24	712-714/5.3	We welcome that the document D4.3.2 provides guidance on the appropriate use of population adjustment methods. However, at various occasions the proposed guidance goes beyond a methodological guide. The sentence *"This means that a conclusion can be drawn regarding an effect only if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect."* corresponds to an instruction on the levels of uncertainty that should be acceptable for decision making. Whether or not a given level of uncertainty is acceptable for a specific decision with respect to a policy question should be left to Member States as part of their appraisal process. We recommend removing this as well as similar sentences in other sections, namely p.5, l.124-126; p.24, l.712-714; p.27, l.801-803.	The use of shifted hypothesis testing is described as option. This sentence gives only an explanation what it means to use the approach of shifted hypothesis testing if it is used. Whether it is used is left to the Member State.
Bayer	24	656-657/ Section 5.2	Lines 656-657 state that the presence of heterogeneity may mask inconsistency in the context of a network meta-analysis. However, inconsistency is itself induced by treatment effect heterogeneity; namely, by imbalances in effect modifiers between the direct and indirect evidence.	Yes, but this kind of heterogeneity is called inconsistency.
Bayer	24	697-698/ Section	Current: When this assumption does not hold, these methods do not yield meaningful results.	Thank you. We have clarified the text.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
		5.2.4	Proposed: When this assumption does not hold, results are accompanied by higher uncertainty and the derived treatment effect may be questioned Rationale: A violation of an assumption does not lead automatically to a useless estimate as suggested in the above statement. Assumptions are not axioms and apodictical within hypothesis logic.	
Bayer	24	708-715/ Section 5.2.4	U,nlike direct comparisons indirect comparisons are not powered a priori for their primary endpoint(s) at least, applying "testing of shifted hypothesis" to derive respective endpoint-specific confidence intervals implying stronger expected effects is a misleading approach in regard of the reduced power of indirect comparisons.	Yes, but this is not relevant, because RCTs are only powered for the primary endpoint. In JCAs we look at a number of other endpoints, for which none of the trials are powered.
James Ryan AZ	24	686-694/Section 5.2.4	The use of a single measure of treatment effect such as the hazard ratio (HR) has considerable benefits, and research teams should not overlook the merits of HR based analyses without serious consideration. Firstly, interpretation of a single treatment effect estimator, such as the HR, is easier to communicate and cross compare to data from individual clinical trials (where the HR is often the standard treatment effect estimator). Furthermore, the advanced methods discussed have often proved challenging to implement. Models often struggle to converge and/or provide implausible results. Specifically, time dependent hazard ratios can put more focus on the results from the tails of a Kaplan Meier curves where there are few patients at risk. As such, we suggest more background is provided. We provide a suggestion of modified text below: **Suggested Text** In cases involving time-to-event data, evidence synthesis is often based on reported hazard ratios. Analyses using reported HRs perform optimally when the proportional hazards assumption is met. Although in the presence of minor deviations from propotionality, the HR can be taken as representing an average of the treatment effect across the trial period [Ref 1], interpretation becomes especially	Thank you for your comment. We have amended the text to highlight that analysis of published HRs can be robust to minor violations of the PH assumption. We have not included the full level of detail suggested in your comment as further details are given in the associated practical guideline (Practical Guideline "Direct and Indirect Comparisons" (deliverable D4.3.1)).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			challenging when estimated survival functions intersect and can have an impact on decisions that are based on comparisons of expected survival. In these cases, NMA based on parametric survival curves [51] or fractional polynomials [39] can be applied, for which the measure of effect is multidimensional as opposed to a single hazard ratio. Other emerging methods for time-varying hazard ratios described in the literature may also be considered [88]. However, although these models are appealing conceptually, they may be both difficult to interpret and use in practice [Ref 2]. Whatever the method used, prerequisites and assumptions related to the method must be clearly specified and justified. **Reference:** Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Statistics in medicine. 2011;30(19):2409-21. Wiksten A, Hawkins N, Piepho H-P, Gsteiger S, Nonproportional Hazards in Network Meta-Analysis: Efficient Strategies for Model Building and Analysis. Value in Health. 2020 23 (7): 918-927	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	24	692 / 5.2.4	<u>Current wording:</u> "Other emerging methods for time-varying hazards ratios described in the literature may also be considered [88]." <u>Suggested rewording:</u> "Other emerging methods for time-varying hazards ratios described in the literature may also be considered [88]**,[89]**"	Thank you for your comment. We have included this reference.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			[note: **bold** denotes proposed inclusion] <u>Rationale:</u> We suggest the addition of a reference (Freeman, 2022)[1] to substantiate this statement. The proposed referenced paper provides details of other methods for undertaking NMA when non-proportional hazards are encountered. These methods include Royston-Parmar cubic spline models, restricted mean survival time, and piecewise exponential models. <u>References:</u> 2. Freeman SC, Cooper NJ, Sutton AJ et al. Challenges of modelling approaches for network meta-analysis of time-to-event outcomes in the presence of non-proportional hazards to aid decision making: application to a melanoma network. Statistical Methods in Medical Research. 2022 (available at: https://doi.org/10.1177/09622802211070253).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	24	697 / 5.3	<u>Current wording:</u> "When this assumption does not hold, these methods do not yield meaningful results." <u>Suggested rewording:</u> "When this assumption does not hold, ~~these methods do not yield meaningful~~ results **are accompanied by higher uncertainty and the derived treatment effect may be questioned."** [note: **bold** and ~~superscript~~ denotes proposed inclusion and deletion, respectively] <u>Rationale:</u> A violation of an assumption does not automatically lead to a meaningless result as suggested in the current statement.	Thank you. We have clarified the text.
Tanja	24	703 / 5.3	**Editorial: Relevant effect modifiers**	Thank you. We have clarified the text.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Podkonjak – Takeda Pharmaceuticals International AG			Current wording: "It is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data." Suggested rewording: "It is important that the relevant effect modifiers and prognostic factory that impact safety that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data. **If this is not given, rationale needs to be provided as to why the identification of one or more relevant effect modifiers has occurred later.**" [note: **bold** denotes proposed inclusion] Rationale: Takeda recommends the inclusion of the proposed statement into the document for instances where this is not fulfilled, or if new effect modifiers have been identified at a later point. For example, emerging data on a biomarker or effect modifiers identified in recent research with a strong biological plausibility. In addition to treatment effect modifiers, Takeda would like to request the sentence also includes prognostic factors which may impact safety or rates / types of observed adverse events.	
James Ryan AZ	24	702-715	Reword: It is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data [45]. In practice, however, one can never be sure that all the relevant effect modifiers are included. The uncertainty that some relevant effect modifiers are not included always remains. Therefore, population-adjusted methods have to be applied with the utmost care. Clear-cut decisions regarding	We do not think these propositions add clarity to the document.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			treatment effects on the basis of population-adjusted indirect comparisons with common comparators are only possible if the size of the estimated effect is so large that this large effect could not be induced by bias due to missing effect modifiers alone. his can be formally be achieved by the testing of a shifted null hypothesis. This means that a conclusion regarding an existing treatment effect can only be drawn if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect. This approach accounts for the uncertainty that some relevant effect modifiers may not be included. To (additions in bold, removal via cross through): **The results from** population-adjusted methods**, just like more simplistic methods, need to be considered within the context of their underlying assumptions and how many of these assumptions have been met, and to what extent. For example, it** is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data [45]. **Clinical validation of the included relevant treatment effect modifiers also help.** In practice, however, one can never be sure that all the relevant effect modifiers are included. The uncertainty that some relevant effect modifiers are not included always remains. **The results from population-adjusted methods need to be interpreted alongside an assessment of their potential bias and robustness, just like assessment of Bayesian methods.** ~~have to be applied with the utmost care. Clear-cut decisions regarding treatment effects on the basis of population-adjusted indirect comparisons with common comparators are only possible if the size of the estimated effect is so large that this large effect could not be induced by bias due to missing effect modifiers alone. his can be formally be achieved by the testing of a shifted null hypothesis. This means that a conclusion regarding an existing treatment effect can only be drawn if the confidence interval lies completely~~	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			~~above or below a certain threshold shifted away from the zero effect. This approach accounts for the uncertainty that some relevant effect modifiers may not be included.~~	
GSK	24	705-715	The position here is that when applying population-adjusted methods for ITCs, only large treatment effects are reliable because we don't know whether there are unknown effect modifiers. In addition to effect modifiers being relatively rare in practice, we also note that the extent to which we have knowledge of effect modifiers is not being taken into account in this guidance. In many situations there is extensive evidence from clinical trials regarding the lack of effect modifiers. Recommend that the likelihood of unknown effect modifiers based on existing data be a consideration when evaluating the results of a pop-adjusted ITC.	In a situation that everything is clear, we do not need a JCA. In the situation that a JCA is required because the benefit of a new treatment is unclear, there are usually also uncertanties regarding potential effect modifiers
Silke Walleser Autiero Medtronic	24	689-692	Fractional polynomials are quite niche, although they can be useful. B-splines can accomplish the same thing and are quite stable.	Thank you for your suggestion. We have included a reference to B-splines as an alternative method for NMA of survival data.
EFSPI	24	703-705	Current wording: "It is important that the relevant effect modifiers that are included are clinically justified and prespecified […]" It is important to provide clear guidance in-text on what is understood by 'relevant'/'clinically justified'. Suggest to also incorporate some of the wording already presented in Key Points IV: Recommended rewording: "It is important that the relevant effect modifiers that are included are clinically justified. External evidence must be provided in support of a variable being an effect modifier, and there should be good evidence a priori that adjustment is likely to reduce bias".	More guidance is provided in the Practical Guideline.
M. Ermisch – GKV-Spitzenverband	24	714-715	In indirect comparisons concerning non-randomised evidence, even if adjusted, we support the idea of a shifted null-hypothesis (with thresholds informed by the actual data quality at stake), provided that suitable guidance is given (s. comment on page 5 line 124).	Thank you.
Silke Walleser	24	711-712	Same comment as the introduction. We suggest stating the null	We think the sentence "This means that

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Autiero Medtronic			hypothesis test here.	a conclusion regarding an existing treatment effect can only be drawn if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect." explain what is the purpose of the test.
James Ryan AZ	24	683-684	NMA allows us to obtain treatment effect estimates for all pairwise comparisons that are consistent with each other: a series of separate direct comparisons do not ensure this, because different studies contribute to different analyses, which then lose face validity. This advantage of NMA over a series of separate direct comparisons is overlooked by the authors and should be incorporated.	No, we do not have overlooked this. This is the reason for the first statement in Section 5: "When treatments have not been directly compared … indirect comparisons are needed."
Richard Birnie and Sarah Smith, Lumanity	24	Section 5.2.4	Discussion of the proportional hazards assumption is important and the inclusion of this paragraph is appreciated. However, it is also a complicated area and more detail is required in this section to provide the reader with sufficient guidance. Conclusions on whether the proportional hazards assumption are not clear yes/no answers (visual inspection of Kaplan-Meier plots and log-cumulative hazard plots can be subjective. For example, if curves cross at the end of the Kaplan-Meier where there are few patients informing the curves, is this a violation of the proportional hazards assumption or are there simply not enough patients contributing to the evidence). Guidance on assessment of the proportional hazards assumption is therefore required. An NMA of parametric curves or fractional polynomials is reasonable when the proportional hazard assumption does not hold. However, for both fractional polynomials and parametric curves the clinical interpretation of the relative effectiveness is not intuitive. Guidance on the pros and cons of these methods and the interpretation of the resulting relative effect estimates is required. Finally, there is an overlap between estimating relative effectiveness	Thank you for your comment. The additional guidance that you recommend will be included in the corresponding Practical Guideline.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			and extrapolation of outcomes. Fractional polynomials may provide a better fit to a time varying hazard function but often give unrealistic extrapolations due to lack of long-term data to support the polynomial function. Additional guidance on the pros and cons of these methods for extrapolation is required.	
James Ryan AZ	24	695	The document would benefit from discussion of an emerging MAIC technique that optimises ESS – e.g. https://doi.org/10.1002/jrsm.1466 In summary: 1. To optimize the ESS with constraints to match the moments. 2. Can produce narrower confidence intervals than the conventional MAIC when the population shows more differences. 3. Can produce a more robust estimation of weights (with less extreme value) than the conventional MAIC, especially when a small sample size. 4. This method provides a model-free framework for weights; analytical results for one covariate can be derived to be linear. **Marginal and conditional effects** Population adjustment methods for indirect comparison are discussed but there is no discussion of marginal versus conditional effects. The authors may want to consider https://doi.org/10.1002/sim.8965 and comments on it https://doi.org/10.1002/sim.8857	Thank you for highlighting this reference. We acknowledge that MAIC methodology is evolving however the purpose of the guideline is not to survey all existing literature on the topic. We do not believe that the first linked reference fundamentally affects the content of the text and therefore it has not been included. Regarding marginal and conditional effects, we have revised the text to clarify the different estimands targeted by each method.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
EFSPI	24	695	The draft guidance should be more explicit about the target population for population adjustment. It is currently only mentioned briefly in Key Points IV but should also be addressed in-text, in Section 5.3. Clarity around the target population is critical for ensuring consistent application and interpretation of population-adjusted effects across JCAs. For example, matching adjusted indirect comparison (MAIC) is unable to provide estimates in a target population other than that of the aggregate data. As discussed by Philippo et al 2017 (DOI: 10.1177/0272989X17725740), this could lead to a situation where two competitors performing MAIC analyses with their own IPD data and the competitor's aggregate data may end up with apparently conflicting conclusions - not because of disagreement about methodology, but because of differences in the target population. Recommended addition: "The target population for the population adjustment must be stated and estimates should be provided for that population".	Thank you for your comment. Additional text has been added.
M. Ermisch – GKV-Spitzenverband	24	705	Prespecification is clearly necessary to avoid "data driven" analysis. We wonder, why this aspect is mentioned that late in the document and only in this context, as it is applicable for any analyses relying on data that are not prospectively gathered. In addition, it should be outlined that analyses using already existing data (e. g. form claims, health records) always present the possibility of bias resulting from possible foreknowledge of data (or "preparatory analysis").	We do not think the guideline needs to state in general pre-specification is necessary as it is general good statistical principle. We have just outlined a specific issue here to consider.
Richard Birnie, Lumanity	24	Line 714	Important guidance on demonstrating that a relative effect is large enough to not be induced by bias is provided. The guidance proposes testing a shifted null hypothesis to address. However, further detail is required to provide guidance on how to construct this hypothesis, i.e. how to select a suitable threshold. What magnitude of difference is required for a decision maker to conclude the relative effect is not due to bias is a highly subjective question. Different people could	Thank you; however, this approach is only described as option. Whether this approach is used and if, which thresholds are applied is left to the member states.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			have different, equally reasonable, views of this. We recognise that definitive rules which will apply to all situations are neither achievable not desirable. However, some more detailed proposals on how to approach the problem and how to define an appropriate threshold would be extremely helpful in making this guidance more practically applicable.	
Matias Olsen, EUCOPE	24,27	710-714 798-803	The guideline recommends the use of a large effect without describing the size of the effect. Some authors (IQWiG) recommend an effect size (e. g. relative risk) between 2 – 5 (see Rapid Report A19-43) Add" "…if the size of the effect is so large that the effect could not be induced by bias due to missing confounders or effect modifiers alone. This can be formally achieved by testing of a shifted null hypothesis. This means that a conclusion for an existing treatment effect can only be drawn if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect. **An effect is considered to be large if the risk ratio is at least 2-5.**"	Thank you; however, this approach is only described as option. Whether this approach is used and if, which thresholds are applied is left to the member states. However, we added the reference to the IQWiG report in Section 6.
Mihai Rotaru - EFPIA	25	718 / 5.3	**Editorial: Similarity and consistency** <u>Current wording:</u> "Population-adjusted methods for indirect comparisons are useful in situations in which an NMA is performed but there is some doubt regarding whether the similarity assumption is valid for some effect modifiers." <u>Suggested rewording:</u> "Population-adjusted methods for indirect comparisons are useful in situations in which an NMA is performed but there is some doubt regarding whether the similarity **or consistency** assumption is valid for some effect modifiers." [note: **bold** denotes proposed inclusion]	Thank you for your suggestion. According to the definitions used in this document, a violation of the consistency assumption that arises from differences in observed effect modifiers is by definition also a violation of the similarity assumption. Therefore, we do not believe that the inclusion of 'or consistency' as proposed is needed.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Rationale: EFPIA recommends this inclusion for completeness for either assumption of similarity or consistency.	
Bayer	25	742-743/ Section 5.3.2	"MAIC requires correct specification of the propensity score model to achieve balance for the effect modifiers after weighting". MAIC does not necessarily require correct specification of the "trial assignment" propensity score model, as long as all effect-modifying covariates (anchored scenario), and potential transformations of these covariates, are balanced after weighting. As is the case for other weighting-based methods (Brookhart et al. 2006, Shortreed and Ertefaie 2017), including covariates that are associated to trial assignment (i.e., are imbalanced across studies), but do not modify effects (anchored) or are prognostic of outcome (unanchored), leads to reduced overlap, precision, and efficiency without improving the potential for bias reduction (Phillippo et al. 2018, Nie et al. 2013). • *Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J. and Stürmer, T., 2006. Variable selection for propensity score models. American journal of epidemiology, 163(12), pp.1149-1156.* • *Shortreed, S.M. and Ertefaie, A., 2017. Outcome-adaptive lasso: variable selection for causal inference. Biometrics, 73(4), pp.1111-1122.* • *Phillippo, D.M., Ades, A.E., Dias, S., Palmer, S., Abrams, K.R. and Welton, N.J., 2018. Methods for population-adjusted indirect comparisons in health technology appraisal. Medical Decision Making, 38(2), pp.200-211.* *Nie, L., Zhang, Z., Rubin, D. and Chu, J., 2013. Likelihood reweighting methods to reduce potential bias in noninferiority trials which rely on historical data to make inference. The Annals of Applied Statistics, 7(3), pp.1796-1813.*	Thank you for your comments. We agree that MAIC does not necessarily require correct specification of the trial assignment model and have amended the text accordingly to reflect that. The question of including non-effect modifying variables in anchored MAIC is discussed in the associated practical guideline; we believe that it is beyond the scope of this document.
Tanja Podkonjak – Takeda Pharmaceuticals	25	718 / 5.3	Current wording: "Population-adjusted methods for indirect comparisons are useful in situations in which an NMA is performed but there is some doubt regarding whether the similarity assumption is valid for some effect	Thank you for your suggestion. According to the definitions used in this document, a violation of the consistency assumption that arises from differences in observed effect modifiers is by

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
International AG			modifiers." Suggested rewording: "Population-adjusted methods for indirect comparisons are useful in situations in which an NMA is performed but there is some doubt regarding whether the similarity **or consistency** assumption is valid for some effect modifiers." [note: **bold** denotes proposed inclusion] Rationale: Takeda recommends this inclusion for completeness for either assumption of similarity or consistency.	definition also a violation of the similarity assumption. Therefore, we do not believe that the inclusion of 'or consistency' as proposed is needed.
M. Ermisch – GKV-Spitzenverband	25	744-754	Given the yet novel and neither widely used nor widely assessed (e.g. by simulation studies) general framework of Multi-Level Network Meta-Regression (ML-NMR) and the sparse guidance information in 5.3.3, a conclusive assessment regarding this point is not possible. We doubt that ML-NMR will substantially contribute to the goal of unbiased effect estimation in the decision situation of HTA in the data-situations for which ML-NMR is proposed: Partly IPD for some trials and aggregated data only for the other trials of the network. Even considering that the general approach can lead to more precise and less biased effect estimates compared to "classical" NMA based on aggregate data, the appealing impression that the IPD-based "adjustment toolbox" helps to keep bias under control might lead to inappropriately frequent use of ML-NMR – with potentially biased results – rather than plainly denying the acceptance of the ML-NMR and requesting the missing high quality evidence. One main concern regarding ML-NMR is the premise of accepting the strong and actually implausible assumption of shared effect modifiers, which cannot be tested without IPD. Effect modification can be intervention-specific, qualitatively (present or not present) and quantitatively (in the slope of effect change). The shared effect modifier assumption implies the same status of effect modification,	As described in Section 5.3.3 ML-NMR has some advantages compared to MAIC and STC. All population-adjusted methods can be useful in the situation described in Section 5.3, namely to confirm the results of NMA, if there is some doubt whether the similarity assumption is valid. More advice regarding population-adjusted methods is given in the upcoming Practical Guideline "Direct and Indirect Comparisons" (deliverable D4.3.1).

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			qualitatively and quantitatively for all treatments in the network. The less the proportion of trials with IPD the more speculative and potentially biased are effect estimates particularly for those pairs in the network, for which only indirect evidence (i.e. partly IPD or even only aggregate trial data) is available – a situation which can be anticipated to be a very frequent one (IPD for one or 2 comparisons from the sponsor and several trials on additional interventions from competitors' trials with aggregated data only). Hence, in special situations ML-NMR may be helpful (high proportion of IPD-trials, covering the main interventions of interest in a very narrow and restricted network, consistent effect modifier terms), but often certainty may be overoptimistically assumed.	
EFSPI	25	752-754	Current wording: "The population-adjusted treatment effects can be estimated for any target population with given covariate values [...]" It is not true that population-adjusted treatment effects can be estimated in any population, it depends on additional assumptions (the assumption of shared effect modifiers, see Philippo et al 2017, DOI: 10.1177/0272989X17725740). Recommend to add text: "Under additional identifiability assumptions (shared effect modifier assumption), population-adjusted treatment effects can be estimated for any target population with given covariate values"	Thank you for highlighting this, we have added text to this effect.
Silke Walleser Autiero Medtronic	25	732-733	It is more than this. For non-linear outcomes, STC should not be used.	Thank you for highlighting this. We have added a comment highlighting how STC inappropriately combines conditional and marginal effect estimates in this scenario.
Sarah Smith, Lumanity	25	Section 5.3.3	Suggest adding in more detail on which data can be analysed using ML-NMR. To our knowledge there have been no ML-NMR methods published for time to event outcomes. This is an important class of outcomes for which these methods are not currently applicable as far as we know. In contrast, population adjustment methods such as MAIC or STC, although limited to pairwise comparisons, can be	Thank you for highlighting this, we have added this as a limitation of ML-NMR.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			applied to time to event outcomes. Providing some specific guidance on how to approach population adjustment for time to event outcomes would be beneficial	
Sebastian Werner vfa	25	725	Please change "with IPD available for the AB trial "to "with IPD only available for the AB trial ". Makes it clearer that IPD is only available for the AB trial not for AC trial, only have aggregate data for AC trial	Thank you. We have amended the draft.
James Ryan AZ	25	727 5.3.1 Simulated treatment compariso n	We would recommend that this section is expanded to consider the following. 1. STC methods that do not actually use simulation may lead to aggregation bias when the link function is non-linear, "the mean of predicted probability is not equal to the predicted probability calculated by the means of predictor" as discussed by Ishak et al. in Simulation and Matching-Based Approaches for Indirect Comparison of Treatments. 2. For the anchored case the STC model may also contain the prognostic variables if that improves the model fit as discussed in Appendix D: Worked example of MAIC and STC of the paper from Phillippo et al (https://doi.org/10.1177%2F0272989X17725740). The STC model should contain all effect modifiers and prognostic variables for the unanchored case. There should be more discussion of the types of variables to include when using this and related methods for population adjustment.	1. Thank you for highlighting this, we have included text to this effect. 2. This is discussed in the associated Practical Guideline.
James Ryan AZ	25	734 5.3.2 MAIC	1. Unlike other methods, MAIC requires overlapped covariates' distribution in trials. This should be emphasised. 2. For the anchored case, MAIC only adjust for effect modifiers, but for the unanchored case, the logistic propensity score model should include all effect modifiers and prognostic variables. This should be emphasised.	1. Thank you for your comment. We have included a sentence on trial population. 2. The topic of unanchored MAICs is discussed subsequently in Section 5.3.4 (which has now been moved to Section 6). We have not made any change to 5.3.2 as this subsection only considers the anchored case.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
EFSPI	25	734	MAICs (Matching Adjusted Indirect Comparisons) are limited to providing a comparison that is adjusted to the population of the study for which only aggregate data are available, which may not match the target population for the decision. This is a limitation that should be added.	Thank you for your comment. The suggested text has been added.
James Ryan AZ	25	744 5.3.3 Multilevel network meta-regressio n	1. To make the model estimable, ML-NMR requires any treatment to be involved in at least one trial with IPD available or sufficient many trials with aggregate data available. If treatment is only included in one trial with only aggregate data available, either informative priors need to be used, or we need to make the shared effect modifier assumption. 2. ML-NMR cannot be applied to time to event data yet; more work needs to be done to generalize this model further. 3. Whether or not the shared effect modifier assumption should or will be made should be discussed more when implementing this method, taking account factors such as mechanism of action.	Thank you for your comment, we have added points 1 and 2 to the text. Point 3 will be discussed in the associated Practical Guideline.
Sebastian Werner vfa	26	756 – 765 / II.5.3.4	The statement "However, in almost all practical applications this strong assumption is not feasible. Therefore, evidence syntheses without a common comparator (i.e.,use of a disconnected network) are highly problematic. When treatment effects are estimated from disconnected evidence networks, methods for analysis of nonrandomised data should be used, although these are also problematic and require access to full IPD from the trials included (Section 6)." sounds very absolute. It would be better to encourage appropriate analysis to investigate, whether the assumptions hold. Suggestion for rewording: "However, in many practical applications this strong assumption may not be feasible. Therefore, evidence syntheses without a common comparator (i.e.,use of a disconnected network) may be biased. When treatment effects are estimated from disconnected evidence networks, methods for analysis of	Changes have been proposed within the text.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			nonrandomised data should be considered (Section 6)".	
Sebastian Werner vfa	26	Key Points IV	"To this end, model and covariate selection strategies should be prespecified and based on transparent criteria." How can pre-specification in practice be achieved? Especially when planning and conducting indirect comparisons, results of the studies are already published.	The model and covariate selection strategies should be specified before the evidence synthesis is performed.
Sebastian Werner vfa	26	Key Points IV	"Therefore, population-adjusted methods for indirect comparisons cannot typically produce reliable estimates of treatment effects when applied to disconnected networks." Change to: "Therefore, estimates of population-adjusted methods for indirect comparisons when applied to disconnected networks must be discussed carefully. Results can support other sources of evidence or (in cases no other evidence is available) give a tentative estimation of the effect."	We think that our original statement is correct.
Liebenhoff, BAH	26	Key Points IV	"To this end, model and covariate selection strategies should be prespecified and based on transparent criteria." Especially when planning and conducting indirect comparisons, the study results are already published. How can prespecification then be achieved?	The model and covariate selection strategies should be specified before the evidence synthesis is performed.
Mihai Rotaru - EFPIA	26	764 / 5.3.4	**ITC – Comparisons with nonrandomised evidence and access to IPD** <u>Current wording:</u> "Comparisons based on nonrandomised evidence require access to full IPD information." <u>Suggested rewording:</u> "Comparisons based on nonrandomised evidence require access ~~to full IPD information~~ **to full IPD for at least one trial**." [Note: **bold** and strikethrough denotes proposed inclusion and deletion, respectively]	Section 5.3.4 deals only with the situation of a disconnected network We moved the content of section 5.3.4 to Section 6 to make this clear.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			Rationale: EFPIA wishes to highlight that HTDs will seldom, if ever, have access to IPD data for all the trials in a network. While this is the ideal situation, population adjusted methods do not require full IPD information; they can be undertaken if IPD are available for at least one trial. This requirement will be met as HTDs will have IPD for their own trial(s). As such, EFPIA requests the amendment to this current statement.	
Roche	26	760-762/5.3.4	The proposed guidance D4.3.2 provides an appropriate and succinct summary of the assumptions made by population-adjusted methods in comparisons of single-arm trials. Whilst we agree that these assumptions can be difficult to meet, we think the guidance should not discourage considering such methods upfront by saying that *"in almost all practical applications this strong assumption is not feasible"*. Disconnected evidence networks will arise in practice (in particular considering rare cancers, ATMPs, orphan medicinal products). Therefore, we recommend replacing the sentences in lines 760-762 with *"These assumptions can be difficult to meet and, therefore, connected networks should be preferred whenever possible"*.	See general response
Roche	26	763-765/5.3.4	The proposed guidance D4.3.2 reflects a preference for methods for analysis of non-randomised data over population-adjusted methods such as unanchored STC and MAIC, stating that the former *"should be used"*. D4.3.2 correctly points out the limitation that these preferred methods require access to full IPD. In practice, manufacturers will have access to the IPD from their own trials, but not to IPD from competitor trials. Therefore, we recommend reformulating lines 763-765 to *"When treatment effects are estimated from disconnected evidence networks and IPD are available from all studies, methods for analysis of non-randomised data should be used. If the data consist of a mix of IPD and AD,*	Section 5.3.4 deals only with the situation of a disconnected network We moved the content of section 5.3.4 to Section 6 to make this clear.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			population-adjusted methods can be considered".	
Tanja Podkonjak – Takeda Pharmaceuticals International	26	764 / 5.3.4	Current wording: "Comparisons based on nonrandomised evidence require access to full IPD information." Suggested rewording: "Comparisons based on nonrandomised evidence require access ~~to full IPD information~~ **to full IPD for at least one study**." [Note: **bold** and strikethrough denotes proposed inclusion and deletion, respectively] Rationale: HTDs will seldom, if ever, have access to IPD data for all the trials in a network due to commercial sensitivity, accessibility and data privacy or ethical considerations in the context of observational data. While this is the ideal situation, population adjusted methods do not require full IPD information; they can be undertaken if IPD are available for at least one trial. This requirement will be met as HTDs will have IPD for their own trial(s). Following existing methods, it is possible to conducted nonrandomised comparisons with IPD data available from one study (the intervention) and therefore we request this statement be revised as recommended.	Section 5.3.4 deals only with the situation of a disconnected network We moved the content of section 5.3.4 to Section 6 to make this clear.
James Ryan AZ	26	760-761	The draft guideline states "However, in almost all practical applications this strong assumption is not feasible." It is important that consideration of the contextual factors are take into account. Disconnected networks are going to be needed for the purpose of comparability in specific circumstances, particular in rare diseases. We recommended to replace this sentence with a more open	Thank you – as this is a methods paper, we believe that such a comment would be out of scope. See also the response to the comment below from Bayer regarding Section 5.3.4 and our general response

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			sentence underlining the difficulty of meeting the assumption.	
Mihai Rotaru - EFPIA	26	Key Points IV, bullet 3	**Editorial: Population-adjusted methods** Current wording: "Population-adjusted methods for synthesis of relative effects (i.e., in connected networks of evidence) depend on the assumption that all relevant effect modifiers have been included in the model. Regression-based approaches for population adjustment such as STC and ML-NMR further require correct specification of the outcome regression model." Suggested rewording: "**For anchored comparisons**, population-adjusted methods for synthesis of relative effects (i.e., in connected networks of evidence) depend on the assumption that all relevant effect modifiers have been included in the model. Regression-based approaches for population adjustment such as STC and ML-NMR further require correct specification of the outcome regression model. **In contrast, MAIC requires correct specifications of the propensity score model to achieve balance for the effect modifiers after weighting.**" [note: **bold** denotes proposed insertion] Rationale: EFPIA recommends the insertion of 'anchored comparisons' as it is currently unclear if the statement is referring to anchored or unanchored comparisons, as there are different assumptions underpinning each approach. The same insertion should also be done for the statements on pages 24 and 25. Furthermore, EFPIA also suggests the inclusion of MAIC in the key points since it is an increasingly accepted population-adjusted approach as mentioned in the previous section.	Thank you for your suggestion. We have not included the reference to 'anchored' as this section has been restructured to only include anchored indirect comparisons. We have amended the text to highlight the importance of model specification for MAIC as well.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Tanja Podkonjak – Takeda Pharmaceuticals International AG	26	Key Points IV, bullet 3	Current wording: "Population-adjusted methods for synthesis of relative effects (i.e., in connected networks of evidence) depend on the assumption that all relevant effect modifiers have been included in the model. Regression-based approaches for population adjustment such as STC and ML-NMR further require correct specification of the outcome regression model." Suggested rewording: "**For anchored comparisons**, population-adjusted methods for synthesis of relative effects (i.e., in connected networks of evidence) depend on the assumption that all relevant effect modifiers have been included in the model. Regression-based approaches for population adjustment such as STC and ML-NMR further require correct specification of the outcome regression model. **In contrast, MAIC requires correct specifications of the propensity score model to achieve balance for the effect modifiers after weighting.**" [note: **bold** denotes proposed insertion] Rationale: Takeda recommends the insertion of 'anchored comparisons' as it is currently unclear if the statement is referring to anchored or unanchored comparisons, as there are different assumptions underpinning each approach. The same insertion should also be done for the statements on pages 24 and 25. Furthermore, we also suggest the inclusion of MAIC in the key points since it is an increasingly accepted population-adjusted approach as mentioned in the previous section.	Thank you for your suggestion. We have not included the reference to 'anchored' as this section has been restructured to only include anchored indirect comparisons. We have amended the text to highlight the importance of model specification for MAIC as well.
Bayer	26	Section 5.3.4	The use of unanchored population-adjusted indirect comparisons such as MAIC and STC (the methods in Section 5.3.4) seems to be	The purpose of this paper is to outline the methods and the limitations of such

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			completely discouraged. For instance, in lines 110-114: "if indirect comparisons are required, only (…) indirect comparisons respecting randomization are appropriate, which means that the evidence network has to be connected". Also, in lines 127-129: "In the case of disconnected networks (e.g., single-arm trials) and any situations with nonrandomised data, complete access to the individual patient-level data is required in order to apply methods that can adequately adjust for confounding." Finally, in lines 597-598: "The use of methods for indirect comparison based on aggregated data is not recommended in disconnected networks". However, disconnected networks with single-arm trials and limited access to patient-level data are common, particularly in oncology. What are the alternatives in this scenario?	methods – indeed, unanchored MAIC can be problematic as outlined. Lack of access to particular data should not be a driving factor in describing methods and their limitations in certain scenarios. The alternative is described in Section 6.
M. Ermisch – GKV-Spitzenverband	26	756 ff.	For population adjusted indirect comparisons (MAIC or STC) it is correctly stated that adjustment can only be applied for the measured potential effect modifiers. However, it should also be emphasised that it is crucial to control that the adjusted population represents the actual target population as per PICO. A crosschecking of the adjusted population with the target population is essential when applying these methods. If the adjusted population only matches a subpopulation with respect to the PICO, no conclusions can be drawn for the remaining or the total target population. Currently, these methods are not satisfactorily established and their use cannot be recommended.	Thank you for your comment. Regarding this paragraph, we do think the most important message is the fact the underlying assumption is almost never met in practical situations and therefore these methods should not be used for disconnected networks. We moved the content of Section 5.3.4 to Section 6.
EFSPI	26	805	Propensity score methods are unbiased if the statistical model is correctly specified. We suggest to recommend the use of doubly robust estimation methods to overcome the limitations of the propensity methods and also mention the doubly robust approach in key points V.	It is not possible to describe all methods and approaches used for propensity scores in this Guideline.
GSK	26	756 – 765	The statement "However, in almost all practical applications this strong assumption is not feasible. Therefore, evidence syntheses without a common comparator (i.e.,use of a disconnected network) are highly problematic.When treatment effects are estimated from disconnected evidence networks, methods for analysis of nonrandomised data should be used, although these are also	Unfortunately, it is not possible to proof that the strong assumption of no unmeasured confounding is valid. We think that our original statement is correct.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			problematic and require access to full IPD from the trials included (Section 6)." sounds very absolute. It would be better to **encourage appropriate analysis to investigate, whether the assumptions hold**. Suggestion for rewording: "However, in many practical applications this strong assumption may not be feasible. Therefore, evidence syntheses without a common comparator (i.e.,use of a disconnected network) may be biased. When treatment effects are estimated from disconnected evidence networks, methods for analysis of nonrandomised data should be considered (Section 6)".	
Sebastian Werner vfa	27	785 – 786 / II.6.1	Please include the established abbreviation IPTW (for "inverse probability of treatment weighting")	This term appears only one time; therefore, the abbreviation is not required.
Edwards Lifesciences	27	801-803/ Section 6.1 General considera tions	Difficult to define the "certain threshold" precisely. We agree that ideally all confounders and effect modifiers should be captured to produce a quasi-randomisation. But we believe that if the analysis is well-done, pre-defined, then the treatment effect could be considered as a good estimate (without specific need to testing a shifted null hypothesis).	The use of shifted hypothesis testing is described as an option; other options are therefore possible.
Mihai Rotaru - EFPIA	27	774 / 6.1	**Population-adjusted methods: Access to IPD** Current wording: "If nonrandomised evidence is available only at the aggregated data level, this is not sufficient for reliable estimation of treatment effectiveness." Suggested rewording: "~~If nonrandomised evidence is available only at the aggregated data level, this is not sufficient for reliable estimation of treatment effectiveness~~ **Comparisons based on nonrandomised evidence require access to full IPD for at least one trial. Ideally, full IPD are available for the estimation of the treatment effect in nonrandomised comparisons. In many situations, full IPD**	We don't believe this is an incorrect statement and therefore no change is required.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			data may not be available for all comparators. Only when anchored comparisons are not feasible, for example due to unconnected networks or comparisons involving single-arm trials, may unanchored comparisons be considered. In these instances, it is recommended a thorough description of the limitations of the unanchored population adjusted ITC and steps taken to address them be included." [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] <u>Rationale:</u> Section 5.3.4 and Section 6 state that in a single-arm trial, a MAIC based on HTD IPD data from the intervention clinical trial when compared to aggregate comparator data would not be acceptable ("highly problematic"). While this is the most robust approach, population-adjusted methods do not require full IPD information and can be undertaken if IPD are available for at least one trial. The methodological guideline should recognise that it will be common that the HTD will not often have access to IPD data from comparators. Therefore, the guideline should provide recommendations or approaches in addressing this situation so that the HTD and assessors can make the most robust comparative effectiveness assessment possible with the available data. Other published methods guidance may be helpful to inform this, where the same limitations of unanchored population adjusted ITC is cited but acknowledges that frequently this is all that can be done given the data that are available and provides guidance on what information should be provided to highlight the potential bias when presenting results in a submission.[1] <u>References:</u> 1. Phillippo DM et al. Methods for population-adjusted indirect	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			comparisons in submissions to NICE. NICE DSU Technical Support Document 18. NICE Decision Support Unit. 2016 (available at: http://www.nicedsu.org.uk/Populationadjusted-ICs-TSD(3026862).htm).	
Mihai Rotaru - EFPIA	27	792 / 6.1	**Comparisons based on nonrandomised evidence – relevant confounders and effect modifiers** Current wording: "In other words, all relevant confounders and effect modifiers have to be included in the model chosen. Again, it is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data." Suggested rewording: "In other words, all relevant confounders and effect modifiers have to be included in the model chosen. Again, it is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data. **If the pre-specification is not fulfilled, rationale needs to be provided why the effect modifiers are used."** [note: **bold** denotes proposed inclusion] Rationale: EFPIA recommends the inclusion of the proposed statement in the document for instances where this is not fulfilled, or if new effect modifiers have been identified at a later point. For example, emerging data on a biomarker or effect modifiers identified in recent research with a strong biological plausibility.	We do not think that this addition is useful in the context of JCA.
Mihai Rotaru - EFPIA	27	796 / 6.1	**Comparisons based on nonrandomised evidence – general considerations** Current wording:	We don't believe that the suggestion wording improves clarity therefore no change is required.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			"Therefore, clear-cut recommendations regarding treatment effects on the basis of indirect comparisons with adjustment for confounding on the basis of IPD are only possible if the size of the estimated treatment effect is so large that the effect could not be induced by bias due to missing confounders or effect modifiers alone. This can be formally achieved by testing of a shifted null hypothesis. This means that a conclusion for an existing treatment effect can only be drawn if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect. This approach accounts for the uncertainty that some relevant confounders or effect modifiers may not be included." Suggested rewording: "**Therefore,** ~~clear-cut recommendations regarding treatment effects on the basis of indirect comparisons with adjustment for confounding on the basis of IPD are only possible if the size of the estimated treatment effect is so large that the effect could not be induced by bias due to missing confounders or effect modifiers alone. This can be formally achieved by testing of a shifted null hypothesis. This means that a conclusion for an existing treatment effect can only be drawn if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect. This approach accounts for the uncertainty that some relevant confounders or effect modifiers may not be included.~~ **if there is still a possibility of impactful bias after adjustment, e.g., due to missing relevant confounders, a thorough investigation of the amount of possible bias should be performed**." [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] Rationale:	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			EFPIA requests the deletion and insertion of the proposed text. As per earlier general EFPIA comments, the methodological guideline may be deemed to implicitly incorporate a value judgement (e.g., shifted null hypothesis) which would be not in accordance with the EU HTA regulation, as stated, '*It is necessary therefore that Union action is limited to those aspects of HTA that relate to the joint clinical assessment of a health technology and to ensure in particular that there are no value judgements in joint clinical assessments in order to respect the responsibilities of Member States pursuant to Article 168(7) TFEU (L 458/3, 14).*" References: 1. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Roche	27	801-803/6.1	We welcome that the document D4.3.2 provides guidance on the appropriate use of nonrandomised evidence. However, at various occasions the proposed guidance goes beyond a methodological guide. The sentence *"This means that a conclusion can be drawn regarding an effect only if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect."* corresponds to an instruction on the levels of uncertainty that should be acceptable for decision making. Whether or not a given level of uncertainty is acceptable for a specific decision with respect to a policy question should be left to Member States as part of their appraisal process. We recommend removing this as well as similar sentences in other sections, namely p.5, l.124-126; p.24, l.712-714; p.27, l.801-803.	The use of shifted hypothesis testing is described as option. This sentence gives only an explanation what it means to use the approach of shifted hypothesis testing if it is used. Whether it is used is left to the Member State.
Bayer	13 & 14	356-359 &	The fundamental assumption of exchangeability is operationalized by assessing the properties of similarity and homogeneity. However, the	Similarity and homogeneity are concepts that can overlap and not totally mutually

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
		369/372/ Section 3.1.1 & 3.1.2	distinction between similarity and homogeneity is unclear. For similarity, it is stated that there must be: "sufficient similarity of all the trials included regarding effect modifiers, which means that there are no differences in the distribution of known and unknown effect modifiers (e.g., sex or age) that modify the true difference between the treatment arms regarding the outcome of interest". Homogeneity is described as follows: "that the relative effectiveness between each pair of treatments is sufficiently homogeneous across all studies comparing those treatments included in an evidence network". According to these definitions, similarity falls into homogeneity, because imbalances in the distribution of effect modifiers lead to the relative effectiveness between a pair of treatments varying across studies. The distinction would be clearer with similarity describing whether there are imbalances in effect modifiers, and heterogeneity describing the strength of the effect modifiers, regardless of balance/similarity.	exclusive. We have tried to distinguish them as best as we could. Both can be the consequence of differences in patient characteristics and/ or designs. We have added a sentence in the homogeneity section for clarity.
Bayer	27	796-803/ Section 6.1	This implies quasi-dramatical effects with respective CIs, which is contradictory taking the induced power losses by indirect comparisons into account. This approach is not state of the art as it relies explicitly on the precision and not primarily on the derived effect estimator. Furthermore, it is not implemented internationally.	No, this do not imply quasi-dramatical effects, only large effects, which is required to account for the uncertainties of nonrandomised evidence. The use of shifted hypothesis testing is an accepted method to ensure formally that a treatment effect exceeds a certain threshold (e.g., in noninferiority trials). This approach is described as option.
Tanja Podkonjak – Takeda Pharmaceuticals International AG	27	774 / 6.1	Current wording: "If nonrandomised evidence is available only at the aggregated data level, this is not sufficient for reliable estimation of treatment effectiveness." Suggested rewording: "~~If nonrandomised evidence is available only at the aggregated data level, this is not sufficient for reliable estimation of treatment effectiveness~~ **Comparisons based on nonrandomised evidence**	We don't believe this is an incorrect statement and therefore no change is required.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			require access to full IPD for at least one study. Ideally, full IPD are available for the estimation of the treatment effect in nonrandomised comparisons. In many situations, full IPD data may not be available for all comparators. Only when anchored comparisons are not feasible, for example due to unconnected networks (i.e. where violation of similarity, homogeneity and consistency exist), or comparisons involving single-arm trials, may unanchored comparisons be considered. In these instances, it is recommended a thorough description of the limitations of the unanchored population adjusted ITC." [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] <u>Rationale:</u> Section 5.3.4 and Section 6 state that in a single-arm trial, a MAIC based on HTD IPD data from the intervention clinical trial when compared to aggregate comparator data would not be acceptable ("highly problematic"). While this is the most robust approach, population-adjusted methods do not require full IPD information and can be undertaken if IPD are available for at least one trial. The methodological guideline should recognise that it will be common that the HTD will not often have access to IPD data from comparators. Therefore, the guideline should provide recommendations or approaches in addressing this situation so that the HTD and assessors can make the most robust comparative effectiveness assessment possible with the available data. We would also like to stress the importance of population adjusted approaches for the interventions which receive marketing authorisations with single-arm trials, common in oncology, ATMPs and orphan medicines. In many situations is not possible to conduct randomised trials and therefore the EU JCA requires a flexible	

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			approach to enable methods to address these situations. Takeda requests the language througout the guigeline be modified to ackowledge the situations when single-arm trials are needed (i.e., RCTs are not ethical, feasible) and that ITCs using IPD from one trial and aggregate data from a comparator may be accepted, as long as the limitations are fully described by the HDT. Other published methods guidance may be helpful to inform this, where the same limitations of unanchored population adjusted ITC is cited but acknowledges that frequently this is all that can be done given the data that are available and provides guidance on what information should be provided to highlight the potential bias when presenting results in a submission.[1] <u>References:</u> 7. Phillippo DM et al. Methods for population-adjusted indirect comparisons in submissions to NICE. NICE DSU Technical Support Document 18. NICE Decision Support Unit. 2016 (available at: http://www.nicedsu.org.uk/Populationadjusted-ICs-TSD(3026862).htm).	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	27	792 / 6.1	<u>Current wording:</u> "In other words, all relevant confounders and effect modifiers have to be included in the model chosen. Again, it is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data." <u>Suggested rewording:</u> "In other words, all relevant confounders and effect modifiers have to be included in the model chosen. Again, it is important that the relevant effect modifiers that are included are clinically justified and prespecified in a statistical analysis plan before analysis of the data. **If the pre-specification is not fulfilled, rationale needs to be**	We do not think that this addition is useful in the context of JCA.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			provided why the effect modifiers are used." [note: **bold** denotes proposed inclusion] Rationale: We recommend the inclusion of the proposed statement in the document for instances where this is not fulfilled, or if new effect modifiers have been identified at a later point. For example, emerging data on a biomarker or effect modifiers identified in recent research with a strong biological plausibility.	
Tanja Podkonjak – Takeda Pharmaceuticals International AG	27	796 / 6.1	Current wording: "Therefore, clear-cut recommendations regarding treatment effects on the basis of indirect comparisons with adjustment for confounding on the basis of IPD are only possible if the size of the estimated treatment effect is so large that the effect could not be induced by bias due to missing confounders or effect modifiers alone. This can be formally achieved by testing of a shifted null hypothesis. This means that a conclusion for an existing treatment effect can only be drawn if the confidence interval lies completely above or below a certain threshold shifted away from the zero effect. This approach accounts for the uncertainty that some relevant confounders or effect modifiers may not be included." Suggested rewording: "**Therefore,** ~~clear-cut recommendations regarding treatment effects on the basis of indirect comparisons with adjustment for confounding on the basis of IPD are only possible if the size of the estimated treatment effect is so large that the effect could not be induced by bias due to missing confounders or effect modifiers alone. This can be formally achieved by testing of a shifted null hypothesis. This means that a conclusion for an existing treatment effect can only be drawn if the confidence interval lies completely above or below a~~	We don't believe that the suggestion wording improves clarity therefore no change is required.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			~~certain threshold shifted away from the zero effect. This approach accounts for the uncertainty that some relevant confounders or effect modifiers may not be included.~~ **if there is still a possibility of impactful bias after adjustment, e.g., due to missing relevant confounders, a thorough investigation of the amount of possible bias should be performed**." [note: ~~strikethrough~~ and **bold** denotes proposed deletion and inclusion, respectively] <u>Rationale:</u> Takeda requests the deletion and insertion of the proposed text. As per earlier general comments, the methodological guideline may be deemed to implicitly incorporate a value judgement (e.g., shifted null hypothesis) which would be not in accordance with the EU HTA regulation, as stated, '*It is necessary therefore that Union action is limited to those aspects of HTA that relate to the joint clinical assessment of a health technology and to ensure in particular that there are no value judgements in joint clinical assessments in order to respect the responsibilities of Member States pursuant to Article 168(7) TFEU (L 458/3, 14).*" <u>References:</u> ▪ Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance).	
Richard Birnie Lumanity	27	Lines 797-804	As per a previous comment, more guidance is needed on specifying a threshold.	The choice of thresholds is left to the member states.
Silke Walleser Autiero Medtronic	27	781-783	Several methods are mentioned but only discuss the use of PSM in detail. All of these methods rely on slightly different assumptions and data requirements. You require IPD for all of them and they are	We agree; however, we decided to describe only the most important method with more detail (see Section

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			different than combining IPD from the trial data discussed above.	6.1). Otherwise, this Guideline would be a textbook describing all epidemiological methods for non-randomised studies, which is out of the scope.
GSK	27	782-783	Which confounding method is preferred as bias removal technique by the different MS?	It is not the intention of this guideline to identify specific preferred methods by MSs.
Sebastian Werner vfa	27	L. 746	Please change "ML-NMR provides a formulation in a more general framework for which full IPD meta-analysis, STC and aggregate NMA can be seen as specific instances "to "ML-NMR provides a formulation in a more general framework with any mixture of IPD and aggregate data for which full IPD meta-analysis, STC and aggregate NMA can be seen as specific instances". Emphasize the implication of ML-NMR with what types of data available.	Thank you for this suggestion, we have added text to this effect.
M. Ermisch – GKV-Spitzenverband	27	779	see comment on p. 27, line 705	See corresponding answer.
GSK	27	802	Any advice on how to determine the "certain threshold"?	No, we don't give a specific advice; this is left to the MSs.
GSK	27	785 – 786	Please include the established abbreviation IPTW (for "inverse probability of treatment weighting")	This term appears only one time; therefore, the abbreviation is not required.
Silke Walleser Autiero Medtronic	28	809-814	The assumptions for PSM usually require 'strong ignorability' which is: 1. Treatment assignment is independent of the potential outcomes, conditional on the observed baseline covariates and 2. Every subject has a non-zero probability to receive either treatment. The first condition is very important and is not discussed here.	We agree; however, the theoretical assumption of "strong ignorability" is included in the statement of Section 6.1 that all methods for nonrandomised evidence with adjustment for confounding require that there are no unmeasured confounders. This statement refers also to propensity scores. We like to keep the guideline as little technical as possible.
Silke Walleser	28	834-835	This is a good point and also holds true for other study designs such	Thank you. Applicability of individual

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
Autiero Medtronic			as RCTs. Does the trial population generalise to the patient population that will likely receive treatment?	studies is described in another practical guideline D4.6 Validity of clinical studies.
EFSPI	28	833	The guideline should provide considerations about the relevance for the JCA of different treatment effects estimated with propensity score methods (ATT/average treatment effect among the treated; ATE/average treatment effect in the entire population)	Thank you for this comment; the issue of ATT vs ATE estimands will be discussed in the Practical Guideline (deliverable 4.3.1).
Bayer	30	860/ Section III Conclusion	"(STC) is a regression-based approach that fits the outcome to an alternative population". The meaning of this sentence is unclear.	Thank you for highlighting this, we have replaced 'fits' with 'extrapolates.'
Bayer	30	871-872/ Section III Conclusion	The ROBINS-I tool is mentioned the first time in the conclusion section. It should be referenced as well as already earlier discussed within the document.	ROBINS-I is already cited in Section 3.2
Richard Birnie Lumanity	30	Lines 874-879	"For some interventions, single-arm or nonrandomised evidence may be the only evidence available for consideration. However, it may well be necessary to deem that this evidence is insufficient for estimation of the relative treatment effectiveness for decision-making." We would recommend that this conclusion needs to be amended to be applicable in practice. Taken together the guidance does not seem to present any achievable scenarios where single arm trials can provide sufficient evidence, even though it is recognised that in some cases this will be the only evidence available. In those cases the alternative may be that decision makers are left with no evidence at all. This could easily result in delayed access to treatments if new RCTs need to be conducted to provide 'sufficient evidence' of relative effectiveness. Alternatively, individual countries may request alternative additional analyses which undermines the primacy of the join clinical assessment as the main source of evidence. At worst some treatments may simply be inaccessible to	Thank you – we don't agree and we consider the comment to be advisory. As this is a methodological guideline it is out of the scope of the guideline to consider access to medicines as this is within the scope of the member states.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG response
			patients. We would ask the authors to reconsider this conclusion and offer some guidance to facilitate decision making in those cases where single arm studies are the only evidence available.	
EFSPI	30	854	Current wording: "[…] which can influence the reliability of the results" Reliability is not the preferred technical term word here. Recommended rewording: "[…] which can increase the risk of bias".	Thank you. We have amended the draft.
M. Ermisch – GKV-Spitzenverband	30	859	Given the methodological weaknesses present for MAIC, the conclusion that accounting for all relevant modifiers in the model is difficult to ensure seems misleading to developers. Thus, the sentence should be changed to "…, is difficult to ensure and, thus, generally not an adequate analysis.	We amended the statement.
Paolo Morgese - ARM	31	877-879	"it may well be necessary to deem that this evidence is insufficient…" While it is understandable that JCA will identify evidence gaps and outstanding uncertainties, it is not acceptable that JCA will deem the data package "insufficient" to estimate relative-effectiveness. The data package is assumed to be sufficient for the EMA review, and the HTAD has likely shared IPD from own clinical trials; deeming the package insufficient will have a overly negative effect on access. In case of evidence gaps and uncertainty, JCA should estimate relative-effectiveness based on existing evidence and outline ways to fill eventual evidence gaps. For example, through RWE generation plans. Furthermore, it is critically important that the HTA analysts have a strong understanding of the biological mechanisms behind the technologies that they are evaluating to avoid leading to erroneous conclusions which are perverse to the evidence that has been presented. For this reason for complex ATMP technologies, the most experienced staff within the HTA body should be used for the assessment or there is adequate supervision in place.	Thank you – we disagree; of course, a data package should be assessed, but the assessment will clearly outline any limitations associated with the data regardless of it being the only data available. In relation to the experience of assessors this is covered within another deliverable and therefore is outside of the scope of this guideline.
GSK	33		Please add the IQWiG methods paper in the list of references.	Thank you. It has been added.

Please add extra rows as needed.

Comments received outside EU/EEA countries without a direct link to the HTAR

Comments received from organisations outside EU/EEA countries were considered for the revision of the document; however, these are not answered individually by the hands-on group.

Name organisation & abbreviation	Country
College of Pharmaceutical sciences, Dayananda sagar university	India
Delta Hat	UK
ISPOR	USA
PHMR Limited	UK

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
ISPOR	General		It is ISPOR's understanding that the primary purpose of the EUnetHTA document is to describe the methods most commonly used for direct and indirect treatment comparisons to provide guidance to assessors in the context of the EU regulation for joint clinical assessment of health technologies.
ISPOR	General		ISPOR appreciates that the document is a methodological guideline and not a prescription which indirect comparison methods are accepted by HTA decision-makers. One can think of many modifications to the methods described in the EUnetHTA document that may improve the relevance and credibility of an indirect comparison given the evidence at hand relative to the research question of interest. We have seen a lot of methodological development for indirect comparison studies in the past decade, and this is expected to continue for the foreseeable future, especially when more and more studies used for regulatory approval do not follow the standard randomized controlled trial (RCT) design. With this in mind, we like to emphasize a few general points here below that are important for any indirect comparison study, whatever the analytical method, and we would recommend incorporating this information in the guideline document. In addition, we raise a number of specific points.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
ISPOR	General		The purpose of a meta-analysis (MA), network meta-analysis (NMA), or another anchored indirect comparison method for RCTs is to estimate the relative treatment effects between competing interventions to inform decision-making for a specific target population of interest. This means that the study population of each of the individual studies included in the evidence synthesis needs to be representative of the target population of interest, which is the case when there are no differences in effect-modifiers between the study populations and the target population. If this requirement for a *relevant* (N)MA or anchored indirect comparison is met, then there are no differences in patient-related effect modifiers between the different RCTs either, a requirement for a *credible* NMA or anchored indirect comparison. Assessing an (N)MA in the context of the decision problem shows that MA and NMA rely on the same assumptions and illustrates the somewhat irrelevant distinction between the concepts of homogeneity, similarity, and consistency as described in the EUnetHTA document. We may even want to avoid using these terms as these are not used consistently in the evidence synthesis literature anyway. To simplify things: 1) For the findings of an (N)MA to be relevant, there should not be systematic differences in patient-related effect-modifiers between the evidence base and the target population of interest; and 2) for a credible NMA (or anchored indirect comparison) we need a connected network of RCTs without systematic differences in known and unknown effect modifiers (related to patient characteristics, study design characteristics, and contextual factors) between studies.
ISPOR	General		The scientific literature provides relevant papers on NMAs specifically tailored to decision-makers and other consumers of these kinds of studies. One publication that is missing and that we like to highlight is the ISPOR guidance paper on NMA. (Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014;17:157-73). This paper provides a systematic overview of the criteria to assess the relevance and credibility of NMA studies.
ISPOR	General		Increasingly, we are faced with an evidence base where for one or several of the competing interventions of interest there is no RCT available; only single-arm studies. A related challenge is disconnected networks. Unanchored indirect comparison studies rely on the assumption of no differences in effect-modifiers AND prognostic factors between single arm or disconnected studies, which is stronger than the assumption of no differences in effect-modifiers for anchored indirect comparisons. However, the statement by EUnetHTA that indirect comparisons involving single-arm trials and disconnected networks are highly problematic may lead decision-makers to automatically reject such indirect comparisons. This does not serve decision-making well, as the alternative is making decisions based on between-trial comparisons in the absence of an explicit analysis. Furthermore, a requirement that an indirect comparison of disconnected networks is only acceptable when it is based on full

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
			individual patient data (IPD) from all studies included is not in line with the reality of data availability. More often than not researchers performing indirect comparisons to support an HTA submission have only access to IPD for a subset of studies included. Definitely, more research is needed regarding appropriate indirect comparison methods for these kinds of scenarios. However, at this point in time, ISPOR recommends emphasizing that a bespoke and innovative methodological approach to synthesizing a challenging evidence base can still be informative and acceptable as long as it is transparent, adheres to common evidence synthesis principles (e.g. consistency), and maximizes the use of available IPD and benefit of randomization from the RCTs that are available.
ISPOR	General		In the context of the discussion about the credibility of an indirect comparison using studies other than only RCTs (e.g. observational evidence) for decision-making, we like to highlight the potential trade-off between internal bias and external bias. Internal bias relates to suboptimal internal validity (i.e., presence of selection bias, information bias, or confounding bias) in the primary studies included in the evidence synthesis. External bias relates to the "mismatch" between the target population of the decision problem and the study populations of the primary studies. For example, do we prefer an indirect comparison where for one of the RCTs in the network the study population is different from the target population regarding an important effect-modifier (external bias), or an indirect comparison where we replace this study with a non-randomized comparative study with residual confounding that is in exactly the correct population (internal bias)? Both analyses provide suboptimal results for decision-making and it may be unclear which analysis is preferable. ISPOR recommends that EUnetHTA outlines such a potential trade-off in their guideline document, rather than only stating the concerns with indirect comparison studies involving observational evidence.
ISPOR	General		Adding illustrative examples to the document will give greater clarity on the factors the assessors will take into consideration to assess the appropriateness of the method(s) and assumptions the manufacturer has used in their indirect comparisons.
ISPOR	General		Structure of guideline: The document currently acknowledges the presence of different data availability settings, however it is currently structured by the type of analysis. I found this led to confusion as I reviewed, because throughout the document (i.e. Section 5 on population-adjusted comparisons and Section 6 on Non-randomized comparisons) it is unclear if the text refers to connected or disconnected networks and anchored vs unanchored comparisons. To avoid confusion, the guideline would benefit from being restructured based on the data availability of the research question instead of the type of analysis. With this approach, the guideline would outline different data availability situations and provide the suitable comparison methods per situation for the purposes of the JCA.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
ISPOR	General		Observational evidence can be very useful supplementary information as part of any evidence synthesis, including indirect comparisons. The U.S. Agency for Healthcare Research and Quality is just now finishing up an update of its Methods Guidance for use of Non-Randomized Trials in Systematic Reviews. Perhaps that document would be a helpful cross-reference for EuNetHTA
Elaine Stamp PHMR	General		Reporting of the methodology is inconsistent throughout the document. As a minimum, as set out in the objectives, we suggest that there is a clear description of each methodology, an indication of when it is appropriate to use, the assumptions and the strengths and weaknesses.
Elaine Stamp PHMR	General		The conclusion states that the guideline will direct assessors towards the pathway that will provide the best estimate of the relative effectiveness with the least uncertainty. The document is also relevant for other stakeholders including those submitting evidence. It would be useful for the document to include decision criteria and/or a schematic, for arriving at the most appropriate methodology. It should also direct the reader to country specific guidelines if this document is not recommending these methods for all countries in the EUnetHTA.
ISPOR	24	686-694	Other methods for time-to-event data include: Royston- Parmar cubic spline models, restricted mean survival time, piecewise exponential models (Freeman et al. Stat Meth Med Res 2022), and the two-step parametric NMA approach introduced by Cope et al. (Res Synth Methods, 2020) This openness to emerging methods should apply in general across all NMA, not just time-to-event data. Methods are constantly evolving and acknowledgement of this should be added to the summary and/or conclusions sections.
Dr K V Ramanath, Dayananda Sagar university	General	421	'RoB' word need to be expand
ISPOR	10-12	Section 2: 263-325	It would be useful for EUnetHTA to discuss the role of RWE in JCA. RWE has become important for demonstrating effectiveness in the real-world setting, particularly to assess effectiveness in subpopulations, inform historical controls and address uncertainty; greater consideration of RWE is certainly an initiative of EMA. However, the guidance text states that observational data could be used only if IPD are available to allow for rigorous adjustment

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
			for confounding. Access to IPD from observational studies and RWE (i.e. registries) may not always be feasible nor ethical due to data privacy considerations, particularly in rare disease settings. Even in an NMA or other type of indirect comparison is done using data from RCTs, only variables reported in the studies can be used for adjustment for confounding, or sometimes the only available evidence for a comparator is from a single-arm trial. Is the Methodological Guideline implying that these data from published observational studies are never going to be considered as part of the evidence for indirect comparisons? What are the options if there is no evidence for a comparator in a PICO except for observational data or single-arm trials? What if the clinical trial for the intervention in the PICO is a single-arm trial because it was conducted in an area of high unmet need for example? Furthermore, there is also no discussion or guidance on the relevance of the locality of the RWE when/if it is used; should a SLR of non-randomized evidence be conducted? Lines 310-312 - use of "Problematic" to describe evidence networks: Describing methods used to connect disconnected networks as "highly problematic" simply because there is no gold standard does not seem appropriate. Instances of disconnected networks are likely to be a frequent issue, particularly given a potentially large set of comparators of interest to cover the 27 member states and the anticipated evidence packages for many ATMPs and orphan and oncology medicines. In these disease states often is it not ethical or feasible to conduct an RCT or, as in the case of oncology, many variations of standards of care exist which may have limited data. Furthermore, describing these analyses and types of evidence as 'highly problematic' is subjective and a value judgement which is outside of the scope of the Regulation. This language stands to potentially bias the future JCA assessors towards only one type of evidence (RCTs) and lead them to disregard or not consider other evidence sources – both are not appropriate as the totality of evidence should be considered for a thorough assessment. It certainly should be acknowledged that non-randomized evidence and disconnected networks have more limitations than RCTs and require certain assumptions (which could be demonstrated to hold), however, we suggest that the Methodological Guidelines softens the tone against non-randomized or disconnected evidence while still listing their required assumptions and limitations. More generally terms such as "highly problematic" (e.g. lines 298, 762-764), "controversial" (e.g. lines 318, 867), and "unreliable" (e.g., line 320) are used throughout the document and should be revised per the rationale provided here and in the general comments and Section 5 comments.
ISPOR	4-5	69-146	The summary, perhaps due its brevity, sounds more stringent and less clear in places than the actual text of the guidance. E.g.: "If any of these assumptions is violated, the results of the corresponding evidence synthesis do not

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
			provide a meaningful estimate of treatment effectiveness." However, the subsequent sections discuss methodological approaches that help an analysis be informative even when some basic assumptions may not hold completely. Perhaps substitute "may not" for "do not".
ISPOR	8-9	213-262	It should become clear from the text what is mentioned with terminology such as 'effect modifiers', 'prognostic variables', 'confounders', 'confounding bias', …etc. Important to make a clear distinction and to use examples to explain the differences. The above mentioned ISPOR paper provides definitions and graphical illustration of the concepts (Jansen et al, 2014, ViH)
ISPOR	6	164-181	The tone of the document switches from being guidance in some sections to quite prescriptive in others. We suggest the tone should be reviewed throughout to reflect that objective of providing guidance. Also, ll. 171-181 seem more related to scope than objective.
Elaine Stamp PHMR	6	147	The introduction section could give more clarity on the objective and scope of the draft document. We were unsure what level of detail was required
ISPOR	7	208-210	All of these guidance references are from books, but the sentence refers to "articles". Independently from these books are guidances from journal articles, which are noted subsequently in the document, that should be considered as well. Perhaps the sentence should be "... original articles cited in these texts ... ."
Anthony Hatswell, Delta Hat	Section 5.3.2	General	The works referenced are correct, however we found that if a missing variable is correlated with the variables included, the method is not biased, though uncertainty is increased. This has important implications for suitability given the high degree of correlation in clinical characteristics https://www.valueinhealthjournal.com/article/S1098-3015(20)30147-9/fulltext
ISPOR	13	327-332	Why do authors make a distinction between consistency and similarity? The document could bring more clarity on the terminology. Currently no guidance is provided on best practices to collect KOL input on relevant patient characteristics, prognostic factors and effect modifiers.
ISPOR	16,19	478-479, 574-576	Bayesian methods. An additional difference between frequentist and Bayesian methods that is not mentioned in the document is the differences in interpretation of results (e.g. credible intervals)
Elaine Stamp PHMR	19	567	Continuity corrections for zero counts are mentioned in the direct comparison section. It does not appear in sections covering other methodologies although it may be necessary for these methods too.
ISPOR	21	612-613	Please replace "adjusted indirect treatment comparisons" with "anchored indirect comparisons" throughout the

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
			document as it is less ambiguous. Also, it would be better to define this term as early as possible in the document.
ISPOR	22	Section 5.1 and 5.2	The Bucher method is special case of fixed effects network meta-analysis with only 2 RCTs; no need to present it separately. The method by Lumley is not really used. Consider moving it to an historical appendix.
ISPOR	24	711-712	The testing of shifted hypotheses represents just one of many sensitivity analyses that could be undertaken to assess the robustness of the population adjusted indirect comparisons. It would be more valuable to present a number of clear recommendations with assessing the validity of population adjustment approaches (requiring a multi-faceted approach), to describe different levels of uncertainty in specific contexts and recommended further analyses which can be conducted to further explore the sensitivity of the results due to the uncertainty.
Anthony Hatswell, Delta Hat	24	711	Here the E-value should be signposted, as a technique that allows the importance of a missing variable to be quantified. https://pubmed.ncbi.nlm.nih.gov/28693043/
Elaine Stamp PHMR	24	711	Testing of a shifted null hypothesis was not something the statistics team were aware of. A reference would be useful. Shifting the hypothesis to test "if the size of the estimated effect is so large that this large effect could not be induced by bias due to missing effect modifiers alone" without any suggestions around what constitutes an acceptably "large" effect or an appropriate amount to shift by. Shifting the null hypothesis seems similar in purpose to quantitative bias analysis (QBA). More information is provided in the below reference. Leahy, Sammon, Kent et al.: Unmeasured confounding in non-randomised studies: Quantitative bias analysis in Health Technology Assessment, To appear in Journal of Comparative Effectiveness Research, 2022.
Anthony Hatswell, Delta Hat	26	Key Points IV (no ling number)	The statement "Therefore, population-adjusted methods for indirect comparisons cannot typically produce reliable estimates of treatment effects when applied to disconnected networks." Is not correct, again given that all that is required is any missing variables to be correlated with included variables. A more accurate statement would be something equivalent to stating that indirect comparisons may be uncertain, which should factor in decision making.
ISPOR	26	755-765	Section 5.3.4 mentions issues with using MAIC/STC as population-adjusted methods for comparisons of single arm trials and these being "highly problematic". The same wording is used to describe approaches using observational data requiring IPD for the comparator. The Conclusion section mentions, in reference to using methods for single-arm/disconnected studies etc., "...the certainty of the results provided by these techniques remains controversial." There is a risk that unanchored approaches for indirect comparisons with single arm trials could be dismissed, or not fully considered, based on the wording in the draft guidance.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
			There is published methodological guidance on the use of these approaches and many published examples of using these approaches in the literature. Although, limitations and interpretation of the results need to be considered carefully, and the approaches used should be tailored to the evidence in each case, these are still considered valid approaches for many HTA agencies and international HTA societies.
Elaine Stamp PHMR	26	755	Section on methods in single arm trials is very brief. "Evidence syntheses without a common comparator (i.e., use of a disconnected network) are highly problematic. When treatment effects are estimated from disconnected evidence networks, methods for analysis of nonrandomised data should be used, although these are also problematic and require access to full IPD from the trials included." In practice (rare disease areas etc.) this is a common scenario, though often access to full IPD from comparator trials is not available (so methods for analysis of nonrandomised data are not possible), no 'acceptable' methods for such a scenario are detailed.
ISPOR	26,27	755-792	Population adjusted indirect comparison methods involving single arm trials, disconnected networks, or other non-randomized evidence do not necessarily require full IPD ifor all studies involved. Frequently they can be undertaken if IPD is available for at least one study. It is a common situation as the manufacturer does not often have access to IPD data from comparators, particularly for innovative medicines. Please clarify Please rephrase, as follows: Pairwise population adjusted indirect comparison methods involving single arm trials, disconnected networks, or other non-randomized evidence require access to full IPD for at least one study. Ideally, full IPD information is available for all studies in the analysis. However, in many situations this may not be available. Also, it is useful to add to the document that only when anchored indirect comparisons are not feasible unanchored comparisons can be considered. In these instances it is recommended that a thorough description of the limitations of the unanchored population adjusted indirect comparison is provided and steps taken to address them be included."
ISPOR	27	784-789 (and all of page 28):	A newer method for estimating a propensity score is available - the Covariate Balanced Propensity Score (CBPS; see Imai, JRSS, 2013).
Elaine Stamp PHMR	27	766	If the aim of section 6 is to provide appropriate guidance to assessors of joint clinical assessments on best practices in conducting comparative effectiveness studies using real-world data, then more information would be required. This should include information on target trial design, and the biases it avoids, how confounders should be identified and selected, other sources of bias and recommendation to report a risk of bias tool.

Please add extra rows as needed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording
Elaine Stamp PHMR	27	782	The instrumental variables (IV) method is included in the same list as multiple regression and matching/weighting on propensity scores, when they are based on different assumptions. Regression and the use of propensity scores are based on the selection on observables (no unobserved confounding) assumption, whereas IV is from a class of quasi-experimental methods which allow for unobserved confounding. This should be made clearer. For completeness, other quasi-experimental methods such regression discontinuity designs (RDD) and differences-in-differences (DiD), should be listed. Information on when G-methods are typically used could be briefly included i.e., to adjust for time-varying confounding.
Anthony Hatswell, Delta Hat	25 (Section 5.3.3)	750	The comment regarding advantages in large networks is not applicable in this section, or should be caveated, as the majority of comparisons compared using these methods are 'unanchored MAIC' or cross study comparisons, not large networks

Please add extra rows as needed.